Human-Centred Design for Improving VR Training of Clinical Skills

by

Thomas Joseph Matthews

A thesis submitted in partial fulfilment of the requirements for the degree of

Masters by Research

Centre for Digital Entertainment

Faculty of Media and Communication

Bournemouth University

February 2023

Thomas Joseph Matthews

Human-Centred Design for Improving VR Training of Clinical Skills

Abstract

With the advent of modern VR technology in 2016, its potential for medical simulation and training has been recognized. However, challenges like low user acceptance due to poor usability are frequently found, hampering wide-spread adoption. This research aims to address the usability of VR clinical skills simulations, particularly focusing on interaction design, and proposes improvements for higher learning outcomes and user retention.

A literature review and a usability case study of an off-the-shelf clinical VR training application was conducted, revealing usability concerns and areas requiring improvement. The prevalent issues include difficulties with controls, hardware and the 'gulf of execution' in broader 'possibility space' - issues that extend beyond direct interaction designs. A market analysis further reinforced these findings, showing gaps in interaction affordances, pointing to design patterns and trends that could be improved for better usability and interaction.

The synthesis of these findings indicate that the limitations of novel interaction schemes and understanding of the VR simulation's 'possibility space' affect the knowledge transferability. Given these issues and limitations in current VR clinical training simulations, this study outlines several Human-Centred Design recommendations for improvement, incorporating findings from wider VR design research.

This research's findings seek to facilitate the development of more user-centric VR training applications, ultimately leading to enhanced training of healthcare professionals and improved patient outcomes. The study sets a foundation for future interaction design work, addressing the primary usability issues and limitations in current VR clinical simulations.

Contents

Abstract....3

List of Tables....6

List of Figures....6

Acknowledgements....7

Author's Declaration....7

Chapter 1 - Introduction....8

1.1 Research Background....8

1.2 Aim and Objectives....11

1.3 Research Questions (RQs)....12

Chapter 2 - Literature Review....13

2.1 Background....13

2.2 Research Collection....14

2.3 Research Analysis....15

2.3.1 Related Literature....17

2.4 Discussion....18

Chapter 3 - User Study Methodology....20

3.1 Design....20

3.2 Participants....25

3.3 Apparatus....25

3.5 Procedure....29

Chapter 4 - User Study Results and Analysis....32

4.1 User Population.... **Error! Bookmark not defined.**

4.2 Results....32

4.3 Analysis....35

4.4 Limitations....37

4.5 Discussion....37

Chapter 5 - Market Analysis 40

5.1 Background 41

5.2 Data Collection 42

5.3 Data Analysis 43

5.3.1 Clinical Actions 44

5.3.2 Action Results 46

5.3.3 Interaction Designs 48

5.4 Evaluation 51

Chapter 6 – Discussion and Recommendations 53

6.1 Human-Centred Design Recommendations 54

Chapter 7 - Conclusion 60

7.1 Future Work 62

References 65

Appendix A – Clinical VR Training Literature Review Table 78

Appendix B – NASA-TLX and SUS User Trial Scores 84

Appendix C – User Error Results 85

Appendix D – Market Analysis References 86

Appendix E – Clinical Skills Representation 88

Appendix F – Market Analysis Interaction Matrix 89

Appendix G – Market Products Skills Representation 90

Appendix H – Design Recommendations 91

List of Tables

Table 1 - Search terms for systematic review 14

Table 2 - User error categories 23

Table 3 - Resuscitation VR Data Log Examples 31

Table 4 - User error results 32

Table 5 - NASA-TLX and SUS scores..33

Table 6 - Bivariate Pearson error/usability correlations 36

Table 7 - List of commercial XR organisations and products 44

Table 8 - List of clinical actions..45

Table 9 - Clinical actions results 46

Table 10 - Organisation action results...48

List of Figures

Figure 1 - Sample questions from SUS and NASA-TLX surveys 22

Figure 2 - Resuscitation VR virtual ward and avatars...27

Figure 3 - Selectable objects in Resuscitation VR and pointer controllers........................28

Figure 4 - Pointing Pattern (i3 Simulations) and Hand Selection Pattern (SimX)...........49

Figure 5 - Visual-guide examples (SimX, inciteVR, UbiSim) 50

Acknowledgements

This research was supported through the Centre for Digital Entertainment with EPSRC grant funding *CDE2: EP/L016540/1*.

My sincere thanks go to my supervisors Dr. Feng Tian and Prof. Wen Tang for their invaluable support over these many years of work, as well as to Mike Board and Zoe Leonard of the CDE team for constant motivation and assistance.

Thanks also go to my industry supervisor Tom Dolby at i3 Simulations, who encouraged both me and my work, and who helped build my work skills alongside this research, and to Dr. Todd Chang at Children's Hospital Los Angeles for supporting the user study and guidance on simulation software.

Finally, my most heartfelt thanks to the Penrose Roots team, to my family, and to my friends Helen, Nadim, Bea, Canvas, Tim, Becca, Jennifer, and many more, who all have supported me through a most difficult year, and to whom I owe a massive gratitude.

Author's Declaration

I declare that this dissertation is all my own work and the sources of information and materials in this research have been fully identified and properly referenced.

Some content of Chapter 3 and 4 of this thesis have been previously included in research journal publication "Interaction design for paediatric emergency VR training" (2020) for Virtual Reality & Intelligent Hardware, DOI: 10.1016/j.vrih.2020.07.006.

Images of Resuscitation VR software have been included with permission from i3 Simulations Ltd.

Chapter 1 - Introduction

This chapter will provide an outline of the research project, the research origins, aims and context, as well as a summary of the core research questions and objectives. This project focuses on answering the question "What are the Interaction Design obstacles to implementing and embedding VR into clinical skills simulation training?"

1.1 Research Background

Core to clinical training are clinical skills, which refer to the abilities and knowledge that healthcare professionals need to perform their duties in a patient-centred and evidence-based manner. They include a wide range of competencies, such as assessment, diagnosis, treatment, and management of patients, as well as communication, teamwork, and professionalism. The acquisition of these skills is typically facilitated through a blend of theoretical education, high-fidelity simulation-based training, and hands-on clinical experience under supervision. This approach offers trainees the opportunity to hone their skills in a safe and controlled environment before transitioning to real-world applications.

Traditionally, high-fidelity simulation training in healthcare, which involves the use of simulated scenarios with environments and patients that closely resemble real-life conditions, is conducted utilizing either patient actors or manikin-based simulations (Datta, Upadhyay and Jaideep, 2012). In the case of clinical skills, manikin-based simulations are favoured over patient actors, particularly for emergency or critical medicine, due to the limitations of patient actors in adequately capturing the severity of the cases being simulated (McFetrich, 2006).

In accordance with guidelines set forth by Health Education England (2020), the UK's Core Medical Training program mandates the provision of annual simulation training for all required procedural skills. However, studies have indicated that a three-month interval between simulation sessions is optimal for retraining clinical simulation scenarios (Hsieh *et al.*, 2015) and that more frequent training is required to mitigate burnout and retain personnel (Iliopoulos *et al.*, 2018).

Identified barriers to running high-fidelity manikin-based simulations include lack of trained staff and time, insufficient and difficult to maintain equipment (Al-Ghareeb and Cooper, 2016), and lack of resources, support, and buy-in from senior decision-makers (Ferguson *et al.*, 2020).

Within recent years, however, there is a growing interest in introducing virtual reality (VR) simulations as a supplement or replacement for manikin-based simulations due to these barriers. VR high-fidelity simulations can potentially be more efficient alternatives or supplementary to manikin-based training, with an overall 30% reduction in resource cost (Haerling, 2018).

Since the arrival of off-the-shelf modern VR technology in 2016, there has been significant interest in using VR for medical simulation and training (Virtual Reality Society, 2020). Within the wider educational context, validity studies support using VR, with evidence that it can improve long-term skills gain and training results (Moglia *et al.*, 2016; Vaughan *et al.*, 2016; Vaughan, Gabrys and Dubey, 2016).

The majority of VR simulations in healthcare primarily focus on surgical training (Moglia *et al.*, 2016; Vaughan *et al.*, 2016), potentially as VR aligns well with the procedural nature of surgical skills. However, there is also wide adoption of VR for clinical skills training as explored in this research.

Despite this history, VR still faces challenges that prevent widespread adoption and reduce interest from decision-makers. One challenge is low user acceptance of VR simulations due to low usability scores (Baniasadi, Ayyoubzadeh and Mohammadzadeh, 2020). A wider review of the literature on clinical VR training simulations and the limitations preventing adoption is conducted in Chapter 2.

Given that most users are new to VR systems (Cohen *et al.*, 2018), and the importance of usability for high learning outcomes, it is vital to prioritise intuitive interactions for successful onboarding and user retention.

For the purpose of this research, usability is define as per ISO 9241-11 (2018) as the "extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use". The research focuses further on interaction usability as established with Norman's (2013) Human-Centred Design and its principles.

In this research, VR specifically refers to "full" VR hardware with positional and rotational head and controller tracking (6 degrees of freedom), distinguished from "mobile" or "limited" VR hardware that only offers rotational head and hand tracking (3 degrees of freedom).

This research builds upon a previous study by Chang et al., (2019), examining the VR clinical skills simulation produced by i3 Simulations (2019). This simulation, created with emergency paediatric physicians, immerses trainees in an emergency medicine scenario where they are required to diagnose, intervene, and stabilise the patient. Further details on scenario and interaction design can be found in Chapter 3.2.

Chang et al., (2019) carried out a study comparing stress levels experienced by emergency physicians during both resuscitation events simulated in a VR application and in real life. Their results indicate that the VR simulation creates an 'optimal' stress environment for learners. Such 'optimal' stress could improve training effectiveness in contrast to simulations with lower stress levels.

However, for VR to be viable in clinical training, further work to "mitigate the novelty of VR is needed" (Chang *et al.*, 2019, sec. Future Directions) in order to improve initial usability for first-time users of VR – indicating a research direction required towards standardised VR *interaction design* for medical simulations. However, with the relatively recent acceptance of VR medical training, there is a lack of evidence-based best practices for *interaction design* tailored for this domain.

Some work on this subject within wider VR research exists. The VR *interaction designs* outlined by Jason Jerald (2016, pt. 5) were inspired by Norman's Human-Centred Design (2013) but primarily based on design predating the release of modern off-the-shelf VR hardware (Virtual Reality Society, 2020). Similarly, the recommendations by Alger (2015) focus solely on VR user interfaces and predate the current generation of VR technology.

Existing studies within the current generation of VR technology mainly focus on locomotion design (Boletsis and Cedergren, 2019; Calandra, Lamberti and Migliorini, 2019; Ntokos, 2019) and do not explore *interaction design* beyond previous defined affordances – such as specifically investigating the interaction design for tracking features of newer VR systems such as the Meta Quest 2 (2020) or Valve Index (2019). Additionally, older studies may cite negative usability results that are no longer applicable due to advancements in hardware capabilities (Freina and Ott, 2015; Carruth, 2017; Jensen and Konradsen, 2017).

1.2 Aim and Objectives

To address this gap, this project will evaluate the usability of current generation VR clinical skills simulations, with a focus on *interaction design*, through the following:

- A literature review of clinical training VR projects, identifying usability concerns and discussions.
- A usability case study of the Resuscitation VR application (i3 Simulations, 2019), with findings presented as applicable to other VR clinical simulations.
- A market analysis of commercially available VR clinical training simulations and their design patterns and trends.

This research aims to propose design recommendations, by exploring standardised *interaction design* and current-generation usability improvements, that address the 'gulf of execution' (Norman, 2013) in VR simulations. This term describe the disparity between a user's current objective (i.e., 'provide patient with airway support') and their understanding of the in-system interactions to complete this goal (i.e., 'use pointer on airway tool in environment').

The usability study will be placed in context to research findings in both academic and industry sectors regarding VR usability and *interaction designs* for training to evaluate gaps in knowledge and generate recommendations for design exploration and development in this sector moving forward.

Overall, this thesis is important because it provides a thorough evaluation of VR in clinical skills training and identifies areas for improvement. The findings from this thesis can inform the development of more effective and user-friendly VR training applications, which can ultimately improve the training and preparedness of healthcare professionals and lead to better patient outcomes.

It is hoped that the recommendations made in this research can be implemented in future work for VR high-fidelity simulations to improve usability, increase adoption of VR technology, and meet the demands of high-fidelity training needs identified by the Health Education England National Strategic Vision for simulation and immersive technologies in health and care (Health Education England, 2020).

1.3 Research Questions (RQs)

This project will seek to address three key research questions:

- RQ 1: What is the current state and limitations of virtual reality clinical skills training applications in healthcare?
  - Chapter 2 will explore this question, with a particular focus on emergency clinical training to align with the Resuscitation VR base project. The aim will be to understand the current state of using virtual reality as a simulation alternative within clinical training praxis.
- RQ 2: What are the barriers to usability for virtual reality simulations in clinical training?
  - Chapters 3 and 4 will explore this question with a user study conducted with both expert and novice clinicians using Resuscitation VR in a training context, to collect and analyse user errors and usability issues. The aim will be to correlate and connect noted user errors with usability scores to determine which areas are of greatest importance, and to outline the gaps in usability and *interaction design*.
- RQ 3: What are potential *interaction design* alternatives that can minimise usability issues?
  - Chapters 5 and 6 will explore this question with a market analysis of existing commercially deployed VR clinical training applications to create an interaction matrix that outlines *interaction design* trends and gaps. The objective is to provide recommendations for future work, incorporating the affordances of current-generation VR hardware.

These research questions address the knowledge gap for both academia and industry and aims to produce original insights and direction for further research.

To address this knowledge gap, the following chapter presents a literature review focusing on the efficacy of virtual reality (VR) in clinical skills training. By reviewing the current body of knowledge, this chapter sets the stage for further exploration of VR-based training programs and identifies areas for future investigation.

Chapter 2 - Literature Review

This chapter presents a literature review of research papers examining the efficacy of virtual reality (VR) in clinical skills training. The review aims to provide a comprehensive overview of the current knowledge regarding the use of VR in emergency and clinical skills training, while identifying areas that require further investigation. It starts by discussing the overall advantages and limitations of VR in this context and then explores specific studies that have assessed the effectiveness of VR in various clinical training settings. The chapter concludes with a discussion of the implications of this research for designing and implementing VR-based training programs, as well as potential future work.

This literature review aims to answer Research Question 1: What is the current state and limitations of virtual reality clinical skills training applications in healthcare?

2.1 Background

Virtual reality (VR) is a technology that allows user to interact with and experience a computer-generated environment in a way that feels 'real' and can be used for a wide range of applications, including entertainment, marketing, and, as in this study, training, and education.

The concept of virtual reality has been around for decades, but early VR systems were limited in their capabilities and found predominantly in large installation. It was not until 2016 with the release of the Oculus Rift CV1 and HTC VIVE hardware that off-the-shelf commercially available technology allowed users to move freely and interact within a fully 3D virtual space (Virtual Reality Society, 2020).

Medical simulations are 'virtual' environments that are used to train healthcare professionals in various clinical skills. These simulations can be real, in the case of patient actors or medical manikins, virtual, in the case of two-dimensional computers programs to more sophisticated VR systems, or, more recently, a combination of the two using mixed-reality and/or haptic interfaces.

Medical simulations in some capacities have been used for training since the dawn of medicine, but what can be considered as the 'modern' medical simulation was a purpose-built medical manikin simulator created in the early 1960s (Cooper and Taqueti, 2008). Since then, simulation technology has advanced and the use of VR for medical training has gained popularity in recent years due to advances in VR technology and the increasing demand for more efficient and effective training methods.

Fundamentally, the use of medical simulations for training has several advantages, including the ability to provide a safe and controlled environment for practicing skills, the ability to repeat and practice specific scenarios, and the ability to train in a wide range of conditions and scenarios that may be difficult or impossible to replicate in real life, such as rare or high-risk situations.

However, there are also challenges to using medical simulations, including the need for expensive equipment and the potential for simulated experiences to differ significantly from real-life situations. The introduction of off-the-shelf VR technology and software should theoretically reduce the resource costs of medical simulations and provide closer parity to real-life situations, but due of the novelty of the current technology it is not yet fully known the extent of the impact of VR compared to traditional medical simulations beyond fiscal advantages.

This literature review aims to identify a spread of studies in which modern VR technology has been used to train clinical skills, and outline the findings, limitations, and connotations of these research.

2.2 Research Collection

A systematic review was conducted to examine the current research on the use of virtual reality (VR) in clinical skills training for medical students. To achieve this, a search was conducted using the EBSCOhost database (2022), with the following terms:

Table 1 - Search terms for systematic review

Base Terms	AND	Excluding
"Virtual Reality" "Immers*" "HMD" "HTC" "Oculus"	"Emergency Medic*" "Clinical Skills" "Resuscitation" "Training" "Sim*" "Education" "Curriculum"	"Surgical" "Surgery" "Therapy"

The search criteria included articles published after 2016. This date range was chosen as it is after the release of commercially available VR hardware such as the HTC VIVE and Oculus Rift. The search yielded 671 articles, of which 458 were selected after removing exact duplicates.

Further exclusion criteria were applied, including language (only English articles were included), results (articles without evidenced interventions were excluded), subject matter (articles not related to clinical skills were excluded), and the audience (articles not targeting medical students were excluded). Finally, articles not focused on immersive virtual or extended reality were also excluded, leaving a total of 35 articles for review.

These terms were selected to focus on clinical skills training using immersive VR hardware, as other related areas of healthcare – such as surgical or therapeutic training – would have different simulation needs, expectations and frictions. For example, a VR surgical training review lists visual fidelity and input device precision as key barriers for usability (Lungu *et al.*, 2021), and a therapeutic VR training review lists user attention issues as a barrier (Hartstein *et al.*, 2022). Neither of these usability barriers align with the findings of clinical VR simulation literature, as will be seen in the following sections.

2.3 Research Analysis

The literature review analysed 35 papers on the use of virtual reality (VR) in healthcare training, with the earliest paper dating back to 2018 and the latest paper from 2022. The reviewed papers are sourced from various journals and conferences, with common contributions from BMC Medical Education, Cureus, and Nurse Education Today journals, each containing three papers within this review. The full details of the papers reviewed are available in Appendix A.

Common authors were found to have contributed to the literature on virtual reality (VR) in clinical skills training. Dr. Todd Chang from Children's Hospital Los Angeles has authored four papers analysed, including studies on stress physiology comparisons between real-life and VR-simulated resuscitations. Jonas Schild from Hannover University of Applied Sciences and Arts has authored three papers, focusing on topics such as immersive multi-user VR for emergency simulation training.

The most frequent hardware used in the reviewed projects was the Oculus Rift, with 14 projects using it, followed by the HTC Vive with 11 projects, and the Meta Quest with 5 projects. The remaining hardware used was Windows Mixed Reality with 2 projects, Gear VR with 1 project, and HoloLens with 1 project. This may be attributed to popularity of these hardware but may be biased due to the search criteria and so is only indicative.

The most frequent topic covered in the reviewed projects was resuscitation, with 10 projects focusing on it, followed by cardiopulmonary resuscitation (CPR) training with 5 projects. Other frequent topics included trauma, nursing, diagnosis, paramedic, and advanced cardiac life support (ACLS), each with 3 projects. The remaining topics were basic life support (BLS), airway, communication, and observation. This may be because resuscitation and CPR training are critical skills that healthcare professionals must have, and the use of VR technology may provide a more immersive and realistic training experience for these skills but may have also been biased by the search criteria used and so is only indicative.

Only 37% of the projects reviewed included control and intervention evaluation. This means that most projects did not compare VR training to a traditional or alternative training system. This can make it difficult to determine the effectiveness of VR training as compared to traditional training methods and highlights an area for future work.

57% of the projects evaluated quantitative factors, primarily through measuring trainee learning outcomes through test scores or external validation. However, among the 20 quantitative evaluations, 2 of them (Katz *et al.*, 2020; Lombardo *et al.*, 2022) reported negative evaluations where VR users had a lower test score in VR compared to traditional trainers, where a significant number of users had difficulty with the VR controls for the respective projects that likely affected their evaluation scores. This suggests that VR training may not be as effective as traditional training methods in certain areas, and that there are still issues with VR controls and hardware that need to be addressed.

91% of the projects included evaluations of qualitative factors, with the Technology Acceptance Model (Goodhue and Thompson, 1995), System Usability Scale (Brooke, 1996), and NASA Task Load Index (Hart and Staveland, 1988) being commonly used evaluation frameworks. These frameworks provide insight into users' perceptions and attitudes towards the VR training experience, which can be useful in identifying areas for improvement.

Common trends identified in the reviewed projects included difficulties with VR controls and interactions (47%), difficulties with the VR hardware (25%), and user requests for verbal communication skills to be included in simulation training (14%). Unfortunately, information on the design of VR controls and interactions is limited in this literature, so a detailed design exploration is not possible with these studies.

These difficulties were found to be more prevalent among trainees who were older and had less gaming experience (Wong *et al.*, 2018; Lombardo *et al.*, 2022). This suggests that age and gaming experience may play a role in how easily individuals adapt to and interact with VR controls and hardware, and that younger users and users with more gaming experience may have an easier time adapting to VR controls and hardware and may have higher levels of engagement and satisfaction with VR training.

Furthermore, the trend of users expecting or requesting verbal communication skills included as part of the simulation training, particularly where their usual training scenario would involve teamwork and/or patient liaison, highlights the importance of including realistic communication scenarios in VR training.

2.3.1 Related Literature

Several additional common citations were found within the literature examined in this review, with the following papers each found cited three or more times in this literature.

A number of systematic reviews on the use of VR for education were found to be frequently cited. Radianti *et al.* (2020) examined VR for higher education and indicate a strong interest in immersive VR technologies for educational purposes. Many of the technologies discussed in the reviewed articles are still in experimental stages, primarily focused on performance and usability testing.

Both Kyaw *et al.* (2019) and Bracq, Michinov and Jannin (2019) performed systematic reviews on VR for clinical education specifically, with the latter focusing further on non-technical clinical skills. These analyses found that VR slightly improved post-intervention knowledge scores compared to traditional learning or other types of digital education, and also demonstrated significant improvements in clinicians' cognitive skills (Kyaw *et al.*, 2019), but found few studies have measured the effects of VR simulation on nontechnical skills development – these too were mostly centred on performance and usability testing (Bracq, Michinov and Jannin, 2019).

Unfortunately, none of the above systemic reviews provide the mentioned usability testing outcomes, meaning trends, or correlations between usability scores and trainee performance cannot be measured.

The issue of VR simulation usability is addressed using haptic feedback directly in two studies, the first being a "VR-enhanced mannequin" (Semeraro *et al.*, 2009) that combines proprietary HMD-based VR with tracked VR gloves and an off-the-shelf medical manikin.

This allowed the trainees to be immersive in a virtual emergency ward which still maintaining physical contact with the 'patient' – to check their pulse and perform CPR, for example. The motivation for this project was to bring VR to established manikin-based CPR training, thus 'increasing the involvement and participation of the trainees'.

McGrath *et al.* (2018) approach this from the other direction. In their consensus process at the 2017 Academic Emergency Medicine Consensus Conference, they examine the advantages, limitations, and challenges of implementing virtual simulation – which includes but is not limited to VR simulations. They explain that, despite the benefits of VR simulations, "one of the major challenges associated with creating [virtual simulation]-based procedural simulators is the need to provide users with realistic tactile sensation, or haptic feedback". Some areas of VR-based emergency medicine training may have limited performance or usability scores due to a lack of realistic controls.

2.4 Discussion

To answer Research Question 1 (*What is the current state and limitations of virtual reality clinical skills training applications in healthcare?*): while VR has a lot of potential in healthcare training, there are still areas that need to be addressed such as issues with controls and hardware as difficulty performing complex tasks with VR controls broke immersion and reduced training quality (Schild, Elsenbast and Carbonell, 2021), and findings strongly support that to establish VR as a viable simulation tool, work must be done to improve unintuitive controls for novice users, as per Chang et al. (2019).

This aligns with findings from studies such as Wu et al. (2022) who found students had difficulty with manipulating objects and navigating within VR, which Crosby et al. (2019) suggests can lead to a split attention effect when trainees are unable to focus solely on the virtual environment.

The literature review suggests that VR will find higher acceptance by learners as it becomes more immersive and realistic (Lombardo *et al.*, 2022), which supports this call for increased usability to increase immersion and realism.

Considering the difficulties found with older users too, not only did lack of familiarity with VR lead to lower scores (Katz *et al.*, 2020) and lower engagement (Putnam *et al.*, 2021), but experts found some dissatisfaction with VR systems as it does not support their "individual quirks" (Liyanage *et al.*, 2019). This suggests a wider issue with the possibility space of VR training simulations, which will be explored further in Chapter 4.

Moreover, the literature review suggests that the initial learning curve with the new technology can impact the knowledge transferability to the VR simulator (Harrington *et al.*, 2018) which means that the additional learning required to gain familiarity with VR must be considered as part of the training itself (Kardong-Edgren *et al.*, 2019).

Where users had difficulty with interacting within the virtual environments (Perron *et al.*, 2021), it is suggested that simulations should explore hand-tracking as it more closely matches real-world interaction (Yang and Oh, 2022). This will be considered and discussed in Chapter 4 onwards.

Additionally, the literature review suggested that there is a clear trend of users expecting or requesting verbal communication skills included as part of the simulation training, particularly where their usual training scenario would involve teamwork and/or patient liaison, this will be discussed during the user study evaluation in Chapter 4 and for future work in Chapter 5.

These findings align with the usability concerns that this project aims to address and provide a useful insight into how to improve the usability of VR in healthcare training, by improving *interaction design*, and making VR more intuitive, immersive, and realistic.

In order to gain a deeper understanding of the aspects of interaction design that had the most significant impact on usability, a user study was conducted with the Resuscitation VR project. Study design aligns with common measures and findings found in this literature review.

The following user study includes an evaluation of the NASA Task Load Index (Hart and Staveland, 1988) and System Usability Scale (Brooke, 1996) commonly found in the literature, with the intention to evaluate both quantitative and qualitative factors to improve usability concerns.

The literature review found that resuscitation is the most studied topic (28%) in this VR clinical training set, which aligns with the topic of this study. The most frequent hardware used in the reviewed projects was the Oculus Rift (39%), which is also used in the user study. This study aims to identify key areas for targeted interventions, which would further enhance the usability of the VR simulation.

Chapter 3 - User Study Methodology

Chapter 3 outlines the methodology used in a user study conducted to evaluate a VR clinical training simulation in a clinical curriculum setting. This chapter details the study design, participants, and data collection and analysis methods. The purpose of this user study was to gather feedback from users on the usability and effectiveness of the VR simulation, to inform future design and functionality of interest. The results of this user study as presented and analysed in Chapter 4.

This user study aims to answer Research Question 2: What are the barriers to usability for virtual reality simulations in clinical training?

3.1 Design

Children's Hospital Los Angeles was the organisation who hosted the Chang et al. (2019) pilot study that formed the launch of the Resuscitation VR software (which is localised to US medical protocol and medications). This pilot study aimed to determine whether stress physiology changes in paediatric resuscitation are equivalent between Emergency Department (ED) shifts and in VR simulations. Resuscitation VR scenarios were evaluated by board-certified ED physicians, who had heart rates and salivary cortisol levels recorded during ED shifts, and during VR sessions. The study found less stress physiology in VR than ED, but VR incurred increased workload perception. As such, further research was suggested for "strategies to mitigate the novelty and "foreign" feel of the VR system" (Chang *et al.*, 2019, sec. Future Directions), which is what this user study and wider research aims to impact.

The previous study by Chang et al. (2019) was focused on identifying the stress physiology of trainees using the VR training, with notes on usability coming during the discussion section of the paper. In comparison, the following user study that forms part of this thesis focuses on usability within the VR application directly, and as such has the following parameters examined which were not explored in the previous study:

1. Usability metrics: The previous study focus on stress physiology changes and used the NASA Task Load Index (Hart and Staveland, 1988) for workload perception. In contrast, this study uses NASA-TLX in combination with the System Usability Scale (Brooke, 1996) to assess the usability of the training system alone.
2. User errors: The following study specifically aims to identify user errors during training performances. It employs the Generic Error Modelling System (Reason,

1990) to record and code each error, which are then cross-referenced with usability scores of NASA-TLX and System Usability Scale to understand the impact of error types on perceived workload and usability.

3. Observation protocol: Sessions are observed by a researcher who records users' performances, including errors, through voice and screen recordings. Users follow a Think-Aloud Protocol, vocalizing their decision-making process and understanding of the scenario.
4. Data logs: The following study captures in-simulation data from the Resuscitation VR application directly, which are cross-referenced with voice and screen recordings to extrapolate user errors.

This study measures user errors during simulation sessions, including, but not limited to, those caused by the 'gap of execution'. These measures are in addition to standard usability metrics also used by Chang et al. (2019), which are cross-referenced with user errors to understand areas of greatest usability impact. By taking this approach, we can evaluate the usability of the entire training system independently, without being limited by potential hardware constraints.

To record usability, this study uses the NASA Task Load Index (Hart and Staveland, 1988) rating scale and the System Usability Scale (Brooke, 1996). Users were prompted to complete a pen-and-paper version of both questionnaires after each simulation performance.

The NASA Task Load Index (NASA-TLX), an evidenced scale (Hart, 2006), is designed to is designed to estimate the perceived workload of tasks that users may complete within a system. It is composed of six subscales: Mental, Physical, and Temporal Demands, Frustration, Effort, and Performance. As suggested by the name, it was developed by NASA in their Human Performance Group to evaluate the workload and effects of workload on an individual or team's ability to perform a task or series of tasks. It has been widely used across many relevant domains including medical simulations (Scafà *et al.*, 2020) and virtual reality (Feick *et al.*, 2020), and as seen in Chapter 2. It was chosen to have a metric to associate workload effects of user errors, and the correlations between, but also for consistency with Chang et al. (2019) which also used NASA-TLX as a usability measure. As this study uses the same software and onboarding protocols as Chang et al., the NASA-TLX workload scores would be expected to be similar, and as such if they are replicated then usability and user error findings can be considered to be similar to those that were discussed but not tracked or examined in the Chang et al. study, and so findings can address the "strategies to mitigate the novelty and "foreign" feel of the VR system".

SUS, or System Usability Scale, is another validated questionnaire (Brooke, 2013), consisting of ten Likert subscales to user's subjective usability evaluations of the system (Brooke, 1996). SUS was developed specifically to be an agile, not overwhelming, usability questionnaire for quick analysis of systems by users, and since being academically published has been cited in over 1,200 research studies and described as an "industry standard" (Brooke, 2013). Like NASA-TLX it has been previously used in relevant simulation domains such as medical training and virtual reality (Adams *et al.*, 2019; Ingrassia *et al.*, 2020; Lerner *et al.*, 2020) and as seen in Chapter 2. It provides precise usability scores that facilitate the analysis of user errors. This method was chosen for its speed and reliability in measuring both usability and learnability. Figure 1 below shows some sample NASA-TLX and SUS subscale items.

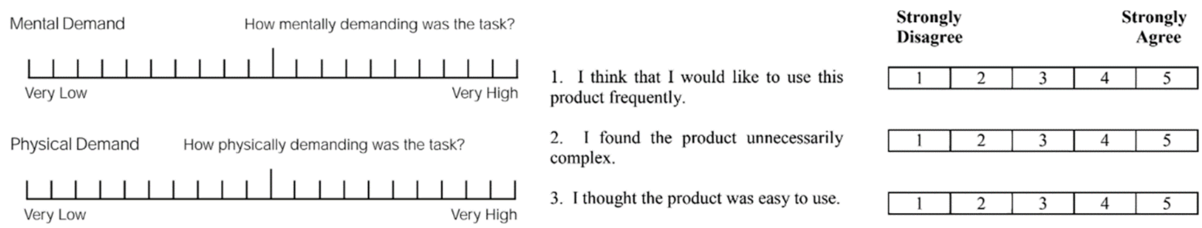


Figure 1 - Sample questions from SUS and NASA-TLX surveys

The Generic Error Modelling System (Reason, 1990) was employed to code user errors while reviewing project recordings, in order to identify where usability gaps resulted in user errors. This is a framework to categorise types of 'error' as occur within a user performance in a system into three types:

- Skill-based errors, usually errors of misplaced attention, control issues, etc.
    - Classified here under controller issues, state misidentifications, and environmental awareness errors.
- Rule-based errors, usually errors of misconstrued understanding of system functionality.
    - Classified in this study under uncoded requests for medications, airway, or procedures, or uncoded action order errors.
- Knowledge-based errors, usually errors related to incorrect domain knowledge.
    - Classified in this study under incorrect usage or medications or airways, or errors following medical protocol. These errors are part of the simulation learning process itself.

For this user study these errors are categorised further as:

Table 2 - User error categories

Skill-Based Errors	
Controller Issue (CI)	Incorrect controller action / inability to interact properly using controllers. Example: Incorrect object selection due to difficulty with using pointer.
State Misidentify (SM)	Misunderstanding of current simulation state / parameters. Example: Orders incorrect medication because they misinterpret visual cues of patient symptoms.
Environment Awareness (EA)	Inability to locate objects within virtual environment. Example: Does not provide airway support because they cannot find the equipment to select.
Rule-Based Errors	
Uncoded Medication Request (UMR)	Requested medication that was not coded in the scenario design and thus not available for this patient. Example: Requests a medication not commonly used for pediatric resuscitation.
Uncoded Airway Request (UAR)	Requested airway tool that was not coded in the scenario design and thus not available for this patient. Example: Requests a type of supplementary oxygen device that is typically used for the patient symptoms.
Uncoded Procedure Request (UPR)	Requested a medical procedure that was not coded in the scenario design and thus not available for this patient. Example: User expects to be able to order specific test unnecessary for the scenario.
Uncoded Action Order (UAO)	Action was requested that required another action to have been completed first before being possible. Example: IV-based medication was requested before any IV line had been established.

Knowledge-Based Errors (Examples)	
Incorrect Medication Usage	Requested a medication that was coded in the scenario design as being unnecessary and/or harmful to the patient. Example: Ordered an adult sedative for a child patient, which would have adverse effects.
Incorrect Airway Usage	Requested a medication that was coded in the scenario design as being unnecessary and/or harmful to the patient. Example: Ordered an oxygen device for a patient with a blocked airway, which would not be effective.
Protocol Error	Failed to follow current medical protocol, as coded in the scenario design for this patient. Example: Ordered an epileptic medication without checking patient blood glucose first, which would indicate which medications should be used.

Skill-Based and Rule-Based Errors were recorded and tracked by the researcher synthesising session recordings, in-simulation data, and Think-Aloud transcripts. Knowledge-Based Errors fall within the purview of the user’s skill with the clinical scenario, and as such, they aren't quantified as being related to usability, they are listed here are examples of errors tracked within the simulation itself. Knowledge-Based Errors are considered to be part of the training outcomes and are attributed to the user's clinical understanding or knowledge rather than being directly linked to the usability of the application.

Due to the nature of Knowledge-Based Errors, it is challenging to determine whether the responsibility lies with the trainee or the training application. Moreover, it is not expected for a training session to not contain Knowledge-Based Errors, unlike Skill/Rule-Based Errors where the aim is their complete elimination from the process. Thus, the presence of Knowledge-Based Errors does not necessarily indicate a flaw in the training application but rather reflects the learning process and the trainee's individual understanding and knowledge acquisition.

3.2 Participants

For parity with this previous study, and as the Resuscitation VR scenarios evaluated were localised to US medical protocols, our study recruited ED physician employees from Children's Hospital Los Angeles in the United States. These physicians were selected because they possess the necessary medical knowledge and expertise to effectively address the resuscitation scenarios presented in the Resuscitation VR application. By targeting users with equivalent knowledge and experience to the intended trainees, our study sought to maintain parity with previous research and ensure the relevance of our findings.

Participants in this study had varying levels of familiarity with VR technology, but none of them had prior direct experience with Resuscitation VR utilized in this research or any other VR clinical training applications. To ensure consistency, individuals who had previous exposure to Resuscitation VR were excluded, including participants in the Chang *et al.*, (2019) study.

In total nine participants enrolled in the study, completing eighteen scenarios in total (not including tutorial scenarios). The potential user population was limited to paediatric emergency physicians at Children's Hospital Los Angeles during a two-week period, with the following criteria:

- Board-certified or board-eligible in paediatric emergency medicine.
- No prior experience with the Resuscitation VR platform or modules.
- Available for 20-minute evaluation session during nonclinical hours.
- No at-risk health factors for use of VR hardware.

A recruitment call was posted on the internal CHLA clinical digital noticeboard and via a copy-all email thread. The nine users who completed this study were the only volunteers who were available within the timeframe and met the eligibility criteria. As this study took place at CHLA itself, clinical emergencies, other simulation sessions, and ongoing clinical work was a factor that prevented more users being available for evaluation.

A benefit of this cohort size however was the ability to engage in detailed post-simulation discussions about the user performance and the software itself, which assisted in identifying User Errors and the underlying stimuli for each.

3.3 Apparatus and Materials

The clinical skills VR simulation evaluated is Resuscitation VR (i3 Simulations, 2019), available on Oculus Rift and Meta Quest hardware, built using Unity game engine (Unity

Technologies, 2018). As the Oculus Rift and Meta Quest versions of the software have identical *interaction design*, and as using the Oculus Rift allowed for greater data tracking as it could have data recorded directly on the host PC, this headset was chosen for testing - this was also consistent with Chang et al. (2019). Other modern VR headsets, like the HTC Vive, would be expected to have comparable outcomes.

Sessions were conducted using the Oculus Rift DK2 headset and two Oculus Touch tracked controllers, on a VR-ready desktop PC. The VR application maintained a consistent frame rate of 90 frames per second. Camtasia, a screen-capture software, recorded the in-simulation sessions, without noticeable effects on the frame rate.

Trainees are immersed in a realistic resuscitation room setting and challenged to complete emergency medicine simulation scenarios. The two scenarios used were paediatric status epilepticus and paediatric anaphylactic shock, which are considered high-risk, low-frequency emergency medicine cases with an Emergency Severity Index of 1-2.

These scenarios were developed in collaboration with pediatric emergency physicians, including authors of Chang *et al.*, (2019), to ensure medical accuracy, and to meet parity with other simulation modalities (i.e. manikin-based simulations) to cover the skills required for the clinical scenario. These scenarios were specifically chosen to address an identified need for high-risk, low-frequency emergency medicine simulations and are appropriate for participants in the test group, who have the appropriate medical training to engage with these scenarios effectively.

The VR scenario replicates the resuscitation room located at Children's Hospital Los Angeles, complete with various medical staff avatars during a resuscitation event. The trainee directs the resuscitation team, assuming the role of 'code captain'. The virtual environment includes the following individuals (avatars), as depicted in Figure 2:

- EMT - Introduces context and outlines initial symptoms of the scenario.
- Nurse - Conducts patient procedures and monitors patient's condition.
- Respiratory Therapist - Conducts airway management procedures.
- Patient – Scenario's focal point, responds realistically to trainee's actions.
- Pharmacist - Administers medication-related interventions.
- Guardian - Emotionally reacts to patient's condition for immersive realism.

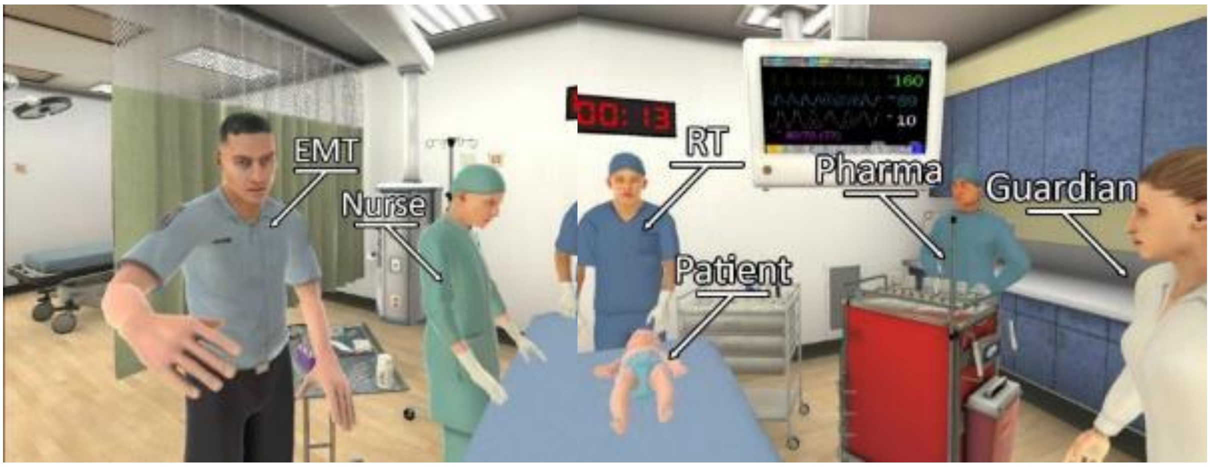


Figure 2 - Resuscitation VR virtual ward and avatars (Images used with permission from i3 Simulations)

The training program involves two scenarios, 'Infant Status Epilepticus' and 'Pediatric Anaphylactic Shock', emulating real-life emergency situations. The scenarios use finite-state machines to simulate patient symptoms, and trainees respond accordingly, following a pre-set, prioritised order of actions, similar to a script concordance test (Fournier, Demeester and Charlin, 2008). Users are not instructed on the medical symptoms prior to starting the simulations, and in-application these scenarios are referred to only as ‘Scenario 1’ and ‘Scenario 2’ prior to session completion.

In the 'Status Epilepticus' simulation, trainees treat an infant experiencing a seizure with a blocked airway, represented through audiovisual cues, virtual agents' communication, and medical examination results (i.e., shallow breaths heard during chest exam). The recommended protocol is: clear airway, provide supplementary oxygen, and administer seizure medication (Lorazepam in this case),

In the ‘Anaphylactic Shock' simulation, trainees treat a child having a severe allergic reaction and difficulty breathing. Like the previous scenario, this is depicted through audiovisual signals and virtual agents' interactions. The recommended protocol is: provide blood pressure medication (Albuterol), administer anaphylaxis protocol (Epinephrine, Methylprednisolone, Diphenhydramine, Ranitidine), and implement advanced airway intubation.

Simulations last about under 5 minutes, with two adjustable aspects: difficulty and distraction. Difficulty modifies the medical complexity of the scenario, and the detail provided in hints from in-simulation avatars. The distraction parameter influences the intensity of various distracting stimuli, such as emotive language and realistic environment noise (like hospital alerts).

For this evaluation, participants completed the scenarios at the Beginner difficulty and Low distraction settings, deviating from Chang *et al.*, (2019) that used Advanced difficulty and High distraction settings in order to invoke increased trainee stress. These settings were chosen to meet current use, as trainees are introduced to the VR simulation on these settings in actual practice. It was also to prevent potential impact on usability scores, as this study is to identify usability issues caused *only* by interaction design, not by scenario difficulty or purposefully induced distractors. Using Advanced difficulty and High distraction settings as per Chang *et al.*, (2019) was concluded to make the correlations between user errors and usability scores less reliable. Finally, the High distraction setting would impact the quality of following Think-Aloud protocol, as the simulation would be more overwhelming and louder, reducing the trainee's capacity to vocalise information.

In this training application, users simulate directing a medical team, using their chosen dominant hand to point and select objects or medications in a Pointing Pattern with semi-realistic hands as per Jerald (2016, sec. 28.1.2). A raytracing technique is employed to identify the nearest interactable object that intersects with the selection ray. Once identified, the object is highlighted (shown in Figure 3), and the user can proceed with the selection by pressing the trigger button on the controller.

The software package Oculus Utilities for the Unity game engine (Unity Technologies, 2018; Oculus, 2020) facilitated these interactions by providing pre-built components such as virtual hands and pointers.

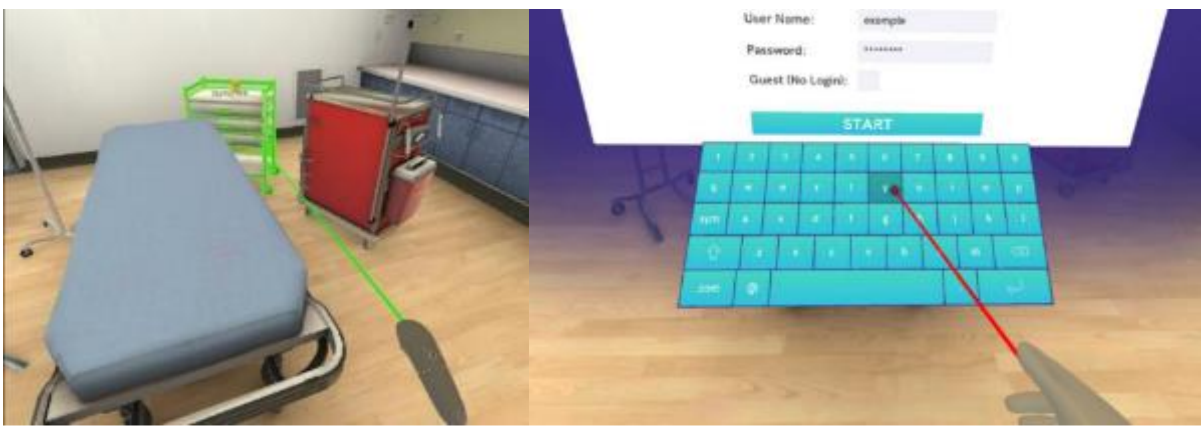


Figure 3 - Selectable objects in Resuscitation VR and pointer controllers (Images used with permission from i3 Simulations)

The selected objects determine the actions performed by the other medical staff present, mirroring real-life resuscitation scenarios. However, some actions and procedures on the patient can be performed by selecting the corresponding hotspot on the patient's body (i.e., checking pulse and capillary refill).

This is a 'standing VR' simulation (Virtual Reality and Augmented Reality Wiki, 2017) in which users cannot physically explore the whole virtual space, but they can pivot, look around, and move within a small area. This design, designed to accommodate Oculus Rift hardware and limited training space at the clinical installation site, requires users to select 'item holders' to then be moved to them via 'blink teleportation' (Ntokos, 2019). At the item holder location, they can then select objects to be used. For example, a user selects the 'Airway Cart' item holder to move to it, and then select an airway tool on the cart itself.

Each selection can prompt four potential outcomes:

- Positive: Correct selection for the scenario, the action will be carried out by the medical team as instructed.
- Neutral: Harmless, yet unnecessary selection that will still be actioned by the medical team as instructed.
- Negative: Incorrect, harmful selection that will prompt appropriate feedback from the medical team and will not be actioned by the team.
- Undefined: Unscripted choice for the scenario, leading to feedback from the medical team and will not be actioned by the team.

The simulation includes a control tutorial that introduces users object interaction and simulation navigation. Scenario-agnostic prompts (such as selecting rubber ducks, for example) are used to ensure users do not inadvertently receive clues about the medical cases prior to starting them.

3.5 Procedure

Prior to simulation sessions, participants were introduced to the study and the simulation, and completed the in-application tutorial, which covered controls and simulation rules. Afterwards participants were introduced to where medical objects and options would be located in the emergency room through physical cue cards (i.e., where to find the medication cart). These cards displayed the types of items they might encounter in each area, without specifying particular objects to avoid priming participants towards certain choices.

The use of cue cards, a practice consistent with Chang et al. (2019), is reflective of the fact that in normal circumstances, users may already be acquainted with the real-world resuscitation room that is presented as the virtual environment in the simulation. This was done to ensure equal footing for all users, as the virtual environment is modelled on the real-world Children's Hospital Los Angeles resuscitation room and thus some users may be more familiar with the physical space than others. This step was taken regardless of trainees' previous experience to mitigate this influence on usability scores.

Participants were also briefed about the expected stimuli load at scenario start, such as the in-media-res debriefing from the EMT transferring the patient to the user's care. This was done to stabilise cognitive load at scenario start as, according to Chang et al. (2019), users who were not familiarised prior to beginning the simulation often failed to hear clinical dialogue given in the scenario introduction. Without this onboarding, the remaining scenario decisions might be adversely affected, thereby skewing usability scores. This kind of onboarding plays a pivotal role in manikin-based simulation, which the VR application supplements or replicates.

Participants proceeded to complete the two resuscitation cases: Status Epilepticus and Anaphylaxis. The same case order was followed for all users to ensure consistent results. Variations in the order of simulations might introduce factors that could impact users' performance or their perception of the scenario and increases the risk of order effects (such as fatigue or practice effects), which could reduce the validity of user scores, particularly in the second scenario.

During sessions, the researcher recorded user errors, while voice and screen recordings of participants' performances were taken. Participants were asked to follow the Think-Aloud Protocol (Dumas, 2001), verbalising their simulation understanding and thought process for each decision and action. This method captured various aspects such as frustrations with controllers, errors noted by the users themselves, and vocalised observations that did not match the then simulation state.

Additionally, data logs were compiled from the simulation software itself, which recorded in-simulation events on a timeline with information tags. Every interaction within the virtual environment as well as every in-application trigger (e.g., a change in patient symptom or an AI character giving a voice line) is recorded as a timestamped entry, with contextual information, in a .csv file which is then parsed by the researcher.

These same data logs are what are used within the application itself to evaluate trainee performance with automated feedback – including Knowledge-Based Errors – and are designed to be comprehensive covering all events within the virtual simulation. Examples of data log entries (formatted for readability) are given as follows in Table 3.

Table 3 - Resuscitation VR Data Log Examples

Count	Action	ActionValue	ActionOutcome	TriggeredTime
1	SCENARIO_STARTED	Seizure_Status_Epilepticus	NULL	15/03/19 16:45
2	SYMPTOM_CHANGED	Seizure State	ACTIVE	15/03/19 16:45
3	SYMPTOM_CHANGED	Vomiting State	ACTIVE	15/03/19 16:45
4	MEDICATION_FAILED	XanaxTabletMedication	NULL	15/03/19 16:46
5	DIALOGUE_PLAYED	NurseAngelCharacter	NA_27	15/03/19 16:46
6	DIALOGUE_PLAYED	NurseAngelCharacter	NA_32	15/03/19 16:46
7	TOOL_USED	NasalTrumpetTool	ACTIVATED	15/03/19 16:46
8	TOOL_USED	NasalTrumpetTool	DEACTIVATED	15/03/19 16:46
9	TOOL_USED	SuctionTool	ACTIVATED	15/03/19 16:46
10	QUANTITATIVE_TIME	TIME_TO_SUCTION	00:50	15/03/19 16:46
11	TOOL_USED	IVLineTool	ACTIVATED	15/03/19 16:46
12	OBJECTIVE_COMPLETED	VomitCureObjective	SUCCESSFUL	15/03/19 16:46
13	TOOL_USED	SuctionTool	DEACTIVATED	15/03/19 16:46

The testing protocol followed was as such:

1. Introduction to the study in a non-VR setting.
2. A tutorial on VR controls and interactions.
3. Onboarding for the VR environment using cue cards.
4. Scenario #1 (Status Epilepticus) in VR.
5. Usability survey and debriefing discussion in non-VR setting.
6. Scenario #2 (Anaphylaxis) in VR
7. Repeated usability survey and debriefing discussion in non-VR setting.

After each session, an unprompted debriefing took place to gather feedback, followed by a structured discussion of the user performance and recorded errors. Full audio from these sessions were later transcribed and synthesised with in-simulation logs and screen-capture videos to further identify Rule-Based and Skill-Based Errors. These steps were integral to ensuring a thorough review of user errors in the system.

In conclusion, Chapter 3 has provided of the key features and interaction design of the VR clinical training simulation, alongside user testing methodology used to evaluate usability and identify causes of user errors. The results of this study, along with an analysis of the findings, are presented in Chapter 4.

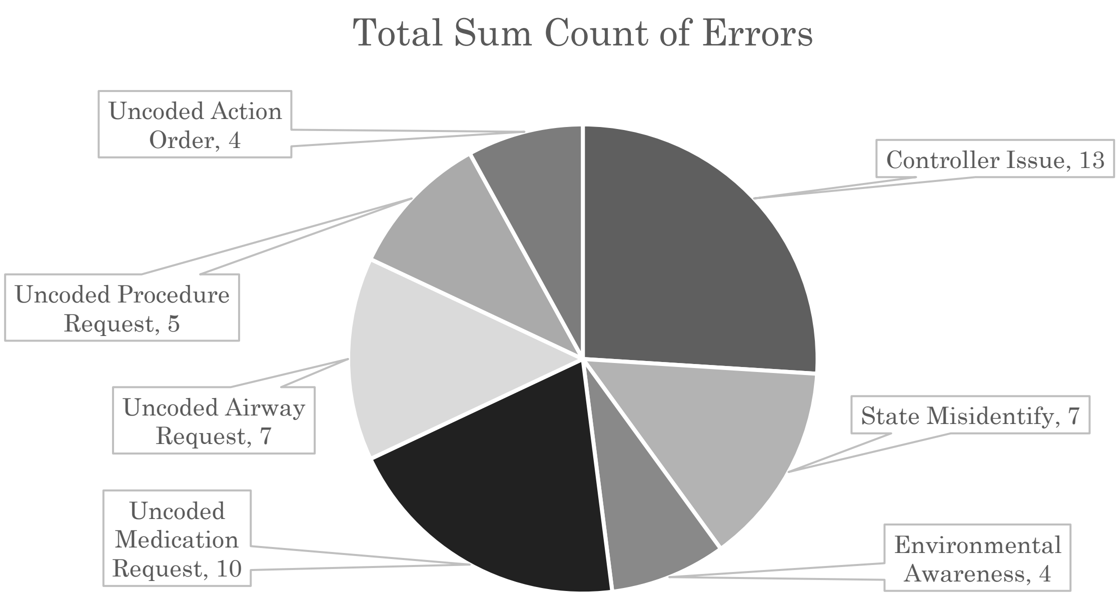
Chapter 4 - User Study Results and Analysis

Chapter 4 presents the results and analysis of the user study conducted to evaluate the VR clinical training simulation as discussed in Chapter 3. The chapter includes a discussion of the results and their implications for the design and functionality of the VR simulation, as well as limitations of the study and suggestions for future research areas.

4.1 Results

The user testing outcomes showed an acceptable usability score, mirroring the Workload scores found in Chang et al. (2019). The only significant category of errors, named "Controller Issue" had a high frequency, with an occurrence average of over 40%. It was found to be directly associated with the interface of the VR simulation, as detailed in Table 4, which provides a breakdown of the total occurrences of each error. Note that a single session can yield multiple instances of the same error.

Table 4 - User error results



A full record of these results, including count (count of sessions with this error recorded, divided by scenario), average per scenario (with this error recorded), and sum per scenario is also available in Appendix C.

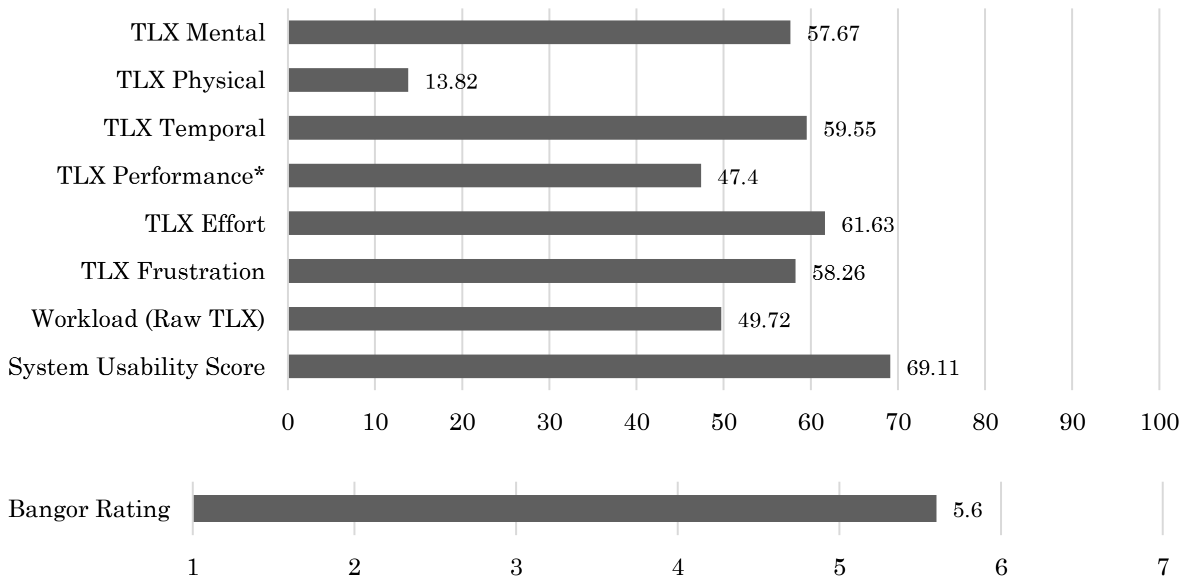
As outlined in Chapter 3.1, NASA Task Load Index (NASA-TLX) and System Usability Scale (SUS) are both validated questionnaires widely used to measure workload and usability. It has been chosen in this study for a proven scale of standard usability metrics and to provide specific scores to correlate user errors against.

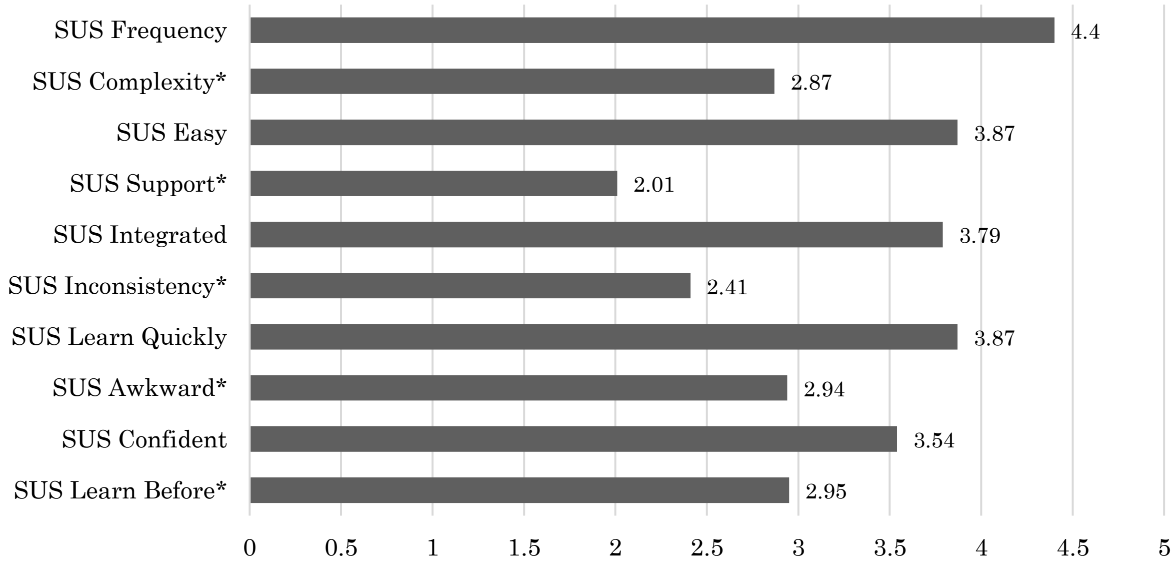
Each subscale on NASA-TLX is scored on a scale from 0 to 100. A higher score generally indicates a more significant task-load, with the exception of the inversed Performance subscale. The overall Workload (Raw TLX) score is an average, with the Performance score inverted to align with the other, negatively scored, TLX scales.

The scoring for each System Usability Scale (SUS) item ranges from 0 to 4, with the usability direction alternating for each factor. For instance, for odd-numbered subscales like SUS Frequency, score is the scale position minus 1, while for even-numbered subscales like SUS Complex, it's 5 minus the scale position. Summing these adjusted subscales together and multiplying by 2.5 to gives the total System Usability Scale score, with a range of 0-100. Alongside these, an overall Banger Rating is given on a 1-7 Likert scale, offering an understanding of users' overall subjective assessment.

The average results are outlined in Table 5. The full scores for each scenario and with statistical deviation are available in Appendix B.

Table 5 - NASA-TLX and SUS scores





*These items are negatively scored so have been inverted for these graphs.

Physical Task-Load recorded a low score, averaging at 13.82 ± 6.76 with remaining TLX scores medium task-load on average. Effort Task-Load was the highest-scoring in this category, with a score of 61.63 ± 13.93. Workload (Raw TLX) also scored as medium task-load, at 47.96 ± 13.11.

For positively-coded SUS subscales – Easy, Frequency, Integrated, Learn Quickly - scores were near Agree on average - Frequency leads with a score of 4.40 ± 0.61. Conversely, negatively-coded SUS subscales – Awkward, Complexity, Inconsistency, Learn Before, Support - showed scores between Neutral and Disagree on average. Support was the highest-scoring negative subscale with a score of 2.99 ± 0.97, with no negative SUS subscale averaging above Neutral.

The tallied average System Usability Score was 69.11, a score that's above the average benchmark percentile with a 'C' grade as per the scale set by Sauro & Lewis (2012), which contextualises SUS scores with the averages across 30 years research findings using the scale. This means that this VR simulation can be considered as above average usability compared to historical system usability scores.

When comparing the scores from the first and second sessions, significant improvements were noted in several areas. There was a +43% improvement in Performance, -11% in Complexity, -9% in Support, -12% in Inconsistency, and -17% in Learn Before. No notable difference was recorded in the overall Workload or the Bangor Rating between the two sessions.

During their scenario sessions, a significant portion of users (61% in total) encountered at least one Controller Issue (CI). Here is a specific breakdown of these issues:

- 38% - Interface Confusion: i.e., users forgetting physical locations of buttons on the controller or unable to aim the controller correctly at objects in the virtual environment.
- 31% - Selection Errors: i.e., making an unintended object selection due to inaccurate pointer targeting.
- 23% - Accidental Presses: i.e., unintentionally cancelling an action due to pressing the wrong button for selection.
- 8% - Assumed Functionality: i.e., trying to give voice commands, assuming the application had this functionality.

In this study, one user's responses proved to be outliers, providing the lowest possible scores on all scales. This deviation exceeded two standard deviations from the mean, and as such these responses were excluded. Despite not officially withdrawing from the study and completing both the NASA-TLX and SUS questionnaires, the participant had shown a strong bias against virtual reality simulations prior to the study. They also ended their VR session prematurely, during the second scenario, which the research team viewed as further evidence of outlier behavior. As such, all the results from this participant were excluded from the final analysis.

4.2 Analysis

The analysis showed a decrease in both the count and sum of errors for most categories when comparing the first and second scenario sessions. The likely explanation for this increase observed in the second scenario can be attributed to users' acclimatisation with the control interface gained during their first scenario playthrough.

However, the category of Uncoded Medication Requests deviated from this pattern, showing a significant 400% increase between sessions. As the anaphylaxis (second) scenario has 6 medications total coded as required or recommended, whereas the seizure (first) scenario has only 1 required medication, this increase may be a false positive because of this greater emphasis on medications in the second scenario.

The highest average error ratio was Uncoded Procedure Requests at 2.5 per scenario. These errors were isolated to a single user, occurring in both their scenario sessions, and could thus be considered as an outlier.

To further examine the relationship between user errors and usability scoring, an analysis was performed with Pearson correlation. The results, shown in Table 6, display only the significant correlations. For example, a strong correlation was found between the occurrence of State Misidentify (SM) errors and negative scores for Complexity, Support, Inconsistency, Learn Before, and Bangor Rating.

Table 6 - Bivariate Pearson error/usability correlations

	TLX Perf.	TLX Eff.	SUS Comp.	SUS Easy	SUS Sup.	SUS Incon.	SUS LQui.	SUS Awk.	SUS Conf.	SUS LBef.	SUS Tot.	Bangor
SM[3]			.568[1]		.514[1]	.515[1]				.555[1]		-.605[1]
EA[4]						-.592[1]	-.518[1]					
UMR[5]	.580[1]		.533[1]	-.533[1]					-.542[1]	.524[1]	-.577[1]	-.695[2]
UPR[6]		.493[1]	.636[1]	-.549[1]				.698[2]	-.541[1]	.660[2]	-.676[2]	-.862[2]

[1]Correlation is significant at the 0.05 level; [2]Correlation is significant at the 0.01 level; [3]State Misidentify; [4]Environment Awareness; [5]Uncoded Medication Request; [6]Uncoded Procedure Request

Interestingly, Table 6, also reveals a strong relationship between Environment Awareness (EA) errors and both positive Inconsistency subscale scores and negative Learn Quickly subscale scores. This may seem counterintuitive as users recorded facing EA errors therefore felt that "there was [not] excessive inconsistency" in the simulation. Given a relatively low occurrence, at 4 total of EA errors, this may be a data outlier.

The category of Uncoded Medication Request (UMR) correlated significantly with negative scoring for subscale Complexity, Easy, Confidence, Learn Before, System Usability, and Bangor Rating, yet had a positive relationship with the Performance subscale. This too appears contradictory, as it implies users committing UMR errors felt more successful in completing their tasks. This inconsistency could be explained by the fact that most UMR errors occurred in the second scenario (anaphylaxis), which also showed a 43% increase in Performance. Hence, it's suggested that this correlation may not result from causation. This ordering of scenarios, which was chosen to align with Chang et al., (2019), could potentially have influenced these correlations, as the second scenario is medically more complex. Future research might assess the impact of scenario order on error correlations.

Lastly, Uncoded Procedure Requests (UPR) showed a significant negative correlation with subscales Effort, Complexity, Easy, Awkwardness, Confidence, Learn Before, System Usability, and Bangor Rating. However, these correlations should be interpreted cautiously as all UPRs were recorded for only a single user.

4.3 Limitations

The findings of this study, due to the limited number of participants involved, have less statistical robustness, and should be viewed as suggestive rather than definitive. Future investigations and user trials that build upon this work would benefit from incorporating a larger participant pool. However, clinical training application evaluation would only be valid with a user population that meets the eligibility criteria of pre-requisite clinical knowledge, as in this study, and as such limited user groups are likely to be a factor. This study population as such was able to provide an unimpeded perspective of the simulation usability as it was designed specifically for their clinical domain and level of knowledge.

In this study, all participants represented the intended user group for the Virtual Reality (VR) clinical simulations. However, to grasp the overall effectiveness of particular design elements and features, an isolated activity independent of previous knowledge would be required. Future work should include such task configurations.

This study, and its emphasis on *interaction design*, doesn't fully encapsulate the overall acceptance of the system. Moving forward, it will be crucial to evaluate the VR simulation's integration within its intended training context in a more comprehensive manner, particularly examining the integration of VR simulation training into medical curriculums as complementary/replacements to manikin-based training, using evaluation tools like the Technology Acceptance Model (Davis, 1989) and Task-Technology Fit (Goodhue and Thompson, 1995) for both trainers and trainees, and evaluating the effectiveness of existing trainee grading, onboarding, and debriefing frameworks applied to VR clinical training. Some extensions of these evaluation tools (Zhang *et al.*, 2017; Bunz, Seibert and Hendrickse, 2021) have been created for VR simulations that could be applied to clinical training as here.

4.5 Discussion

To answer Research Question 2 ('What are the barriers to usability for virtual reality simulations in clinical training?'): the user study results suggest the most significant usability barriers stem from constraints on information exchanged between user and system, rather than issues with existing interaction controls.

Furthermore, analysis of the user errors suggest that the 'gulf of execution' (Norman, 2013) persists not just within direct *interaction design*, but also the broader issue of 'possibility space' (Bogost, 2008). This term, borrowed from game design research, signifies all potential actions permitted by a set of rules (Bogost, 2008, p. 120). Here, it refers to the range and types of user interactions permitted by a system's rules. Identification of this problem and its implications could present a significant contribution to knowledge in VR simulation design.

This challenge of offering sufficient 'possibility space', to fulfil both user expectations and developer limitations, has been discussed in game studies (Jones, 2008) and gameful pedagogies (Caravella, 2019; Cooke, Dusenberry and Robinson, 2020). Some work on 'possibility space' in VR training exists (Miller, Willemsen and Feyen, 2018; Gordon *et al.*, 2019), but these studies, like ours, focus on controller-based *interaction design* as opposed to a holistic evaluation and do not examine the limitations of information exchange.

Despite the high incidence of Controller Issues (CI), it appears they don't significantly impact the system's usability scores. This suggests that users differentiate between the software's usability and any hardware-related problems. A plausible explanation for this could be that CIs interfere with the immersive experience (Jerald, 2016, sec. 4.2), reallocating the mental load away from the training application. While this may influence immersion and learning outcomes, it doesn't seem to affect the usability scores. This merits further exploration in future work.

With the NASA-TLX scores, there is an indication of moderate task-load overall. This reaffirms Chang et al.'s 2019 research, suggesting the simulation maintains 'optimal' stress levels for clinical simulation. Here, 'optimal' stress signifies a mental load (in relation with TLX scores) that doesn't bore or overwhelm the user but keeps them engaged and focused for optimal knowledge retention.

The 'C' grade System Usability score - the 69.11 total graded as per Sauro & Lewis' scale (2012) - indicates that the clinical simulation is appropriate for VR novices, but it also highlights potential areas for improvement. There is correlation between usability and errors related to State Misidentify (SM) and Uncoded Medication Requests (UMR).

A review of other VR clinical simulation research (Schild, Misztal, *et al.*, 2018; Latham *et al.*, 2019) found similar *interaction designs* as in this evaluated simulation, indicating that the 'gulf of execution' may be a widespread friction, and thus this research could have a widespread impact on improving VR simulation usability and addressing the issues affecting uptake as outlined in the literature review.

To validate such 'possibility space', information exchange, and 'gulf of execution' issues, to understand current *interaction design* patterns established to counter these identified frictions, and to evaluate further gaps and future work in usability design, a market analysis was conducted on contemporary VR clinical training simulations and is presented in the following Chapter 5.

Chapter 5 - Market Analysis

Chapter 5 of this thesis, Market Analysis, aims to provide a comprehensive overview of the current market of commercially available virtual reality (VR) medical simulation products and explore *interaction designs*.

This market analysis, and synthesised recommendations in Chapter 6, aims to answer Research Question 3: What are potential *interaction design* alternatives that can minimise usability issues?

As such, this analysis was spurred by the outcomes of the user study in Chapter 4, which prompted a research direction to explore specifically the 'gulf of execution' as a major factor in usability. The focus of this market analysis follows the user study by design: for example, a user study that found Environmental Awareness to have the highest significance on usability would explore environmental and spatial design within VR clinical simulations.

The decision to analyse commercial products rather than software used in research projects, such as those discussed in the literature review, was based on the practicality of their availability for closer examination of *interaction design*. Commercial products, being readily available off-the-shelf, are expected to have undergone numerous iterations and improvements based on feedback from healthcare professionals over the course of the company's history.

This is believed to make them a valuable source of information for this analysis, providing insights into the ways in which VR technology has been integrated into clinical training and what aspects of *interaction design* have proven to be effective for this purpose.

The chapter begins with a background on the healthcare industry's usage and adoption of VR, providing historical examples of its evolution and implementation in medical education and training.

The second section of the chapter presents data collected from online supplier lists of commercially available medical simulation products in VR. This includes a comprehensive list of products currently available on the market, their features, and the targeted areas of training.

The third section of the chapter presents an interaction matrix, where each product is evaluated for its *interaction design*. This includes an analysis of the interaction patterns, input methods, and commonalities in design among the products.

Finally, the chapter concludes with a discussion of the findings and how they support future work in the field of VR medical simulation. The discussion highlights the state of the market, the current trends in *interaction design*, and the potential for further research and development in the field.

5.1 Background

The use of virtual reality technology in healthcare training dates to the 1960s, when Robert Mann developed the first VR system for medical education in 1965 (Pantelidis *et al.*, 2018). This system was designed to create a new training environment for orthopaedic procedures. However, it was not until the late 1980s that the head-mounted display (HMD) was introduced as a wearable device for VR visualizations in medicine (Barteit *et al.*, 2021). The first interactive applications in medical education using this technology were focused on hands-on procedures, and they appeared over a decade later.

The use of VR in medical education and training has been supported by various organizations, including the American Medical Association, which recognized the potential of VR to enhance medical education and improve patient outcomes. The World Health Organization (WHO) predicts that by 2030, the world will require over 40 million new healthcare professionals (World Health Organization, 2022), including doctors, nurses, and frontline healthcare workers, which equates to approximately doubling the current medical workforce. To maintain and grow a sufficient medical workforce, VR is seen as a potential solution to supplement existing training models. WHO highlighted that ineffective training to address the increasing complexity of medical technology and longer learning curves are the primary causes of adverse events related to new technologies.

In recent years, the healthcare industry has seen a significant increase in the adoption and use of VR technology in medical education and training (Blueweave Consulting, 2022). This may be attributed to the availability of affordable and accessible VR technology, as well as the growing body of research supporting the effectiveness of VR in medical training.

The healthcare software and content industry is projected to reach a global market value of $1.72 billion by 2028 (Blueweave Consulting, 2022). Within this industry, training is the third largest sector, following surgery and pain management. However, "lack of adequate training for medical professionals to adopt VR is a significant factor that is anticipated to impede the market growth" (Blueweave Consulting, 2022), in particular due to complex controls and high usability barriers to entry. This demonstrates the need for improved *interaction design* as VR healthcare training products are expected to play a significant role in the industry's growth and development.

5.2 Data Collection

For this scoping review of commercially-available medical training products, examples were collected from HealthySimulation.com (WaterWell LLC, 2022) under the category 'Digital / Virtual Patients', the Society for Simulation in Healthcare (2022) Corporate Roundtable, and XR Ecomap (EcoMap Technologies Inc, 2022) with the filters 'Healthcare' and 'Health and Life Sciences'.

HealthySimulation.com, founded by Lance Baily in 2010, is a leading healthcare simulation resource website that offers a wide range of services and information to healthcare professionals around the world. The website provides the latest news, product demos, vendor connection services, and community sharing. Since its inception, HealthySimulation.com has become a reputable source of information and resources for healthcare professionals in the field of medical simulation, with over 2,000,000 page views of simulation-specific content.

The Society for Simulation in Healthcare (SSH) is an international organization established in 2004 that aims to serve a global community of practice, enhancing the quality of healthcare using simulation. The organization's mission is to improve performance and reduce errors in patient care with simulation-based modalities such as virtual reality and task trainers. SSH membership includes a diverse range of healthcare professionals, researchers, educators, and developers from around the globe.

The XR EcoMap, founded in 2022, is a platform that aims to help people understand the organizations working to grow the Augmented and Virtual Reality industry. The platform is presented by the WXR Fund, Meta, the XR Association, and Qualcomm. It is designed to make navigating the rapidly growing XR ecosystem easier by providing a comprehensive directory of over 1,200 resources and organizations working to develop the industry.

This scoping review of industry examples applied the following criteria for selection: the use of VR as a base format, interactivity (i.e., not 360-video based), availability for purchase, examples of implementation in actual medical organizations, and inclusion of clinical skills (i.e., not surgical simulators). These criteria were established to ensure that the review focuses on commercially available, interactive, and relevant virtual reality-based products that are being used for clinical skills training in medical organizations.

5.3 Data Analysis

In this scoping review of VR clinical training commercial products, a comprehensive search was conducted on three key platforms: HealthySimulation.com, Society for Simulation in Healthcare, and XR EcoMap. The results showed that 43 organizations listed under 'Digital/Virtual Patients' on HealthySimulation.com, of which 17 utilized VR software, with 10 for clinical training. Out of the 35 organizations listed on the Corporate Roundtable of the Society for Simulation in Healthcare, 5 utilized VR software, with 3 focusing on clinical training. In the XR EcoMap, 116 organizations were listed under 'Healthcare' and 'Healthcare and Life Sciences' tags, of which 63 utilized VR software, and 9 of these focused on clinical training.

By removing duplicates, a total of 16 clinical training VR software organisations were identified for analysis. This includes i3 Simulations, the developer of the Resuscitation VR platform which was used in the user evaluation. Resuscitation VR and other i3 Simulations modules will be included in this analysis to expand upon those findings and discussion.

To further refine the list of products, only those with evidence of deployment within a healthcare institution were considered. This was done to ensure that products had evidence of being used by clinicians, as opposed to being untested proof of concepts, and thus would hold validity for examining user design. This was determined through a review of company blogs, case studies, and listed clients. After this filter, 4 products were excluded, leaving a total of 12 products for analysis. These excluded products are all listed as demonstrative showcase projects – i.e., to attract healthcare clients and thus could not demonstrate clinical use.

It is important to note that the results of this scoping review are based on publicly available information and are not exhaustive. However, this review provides a comprehensive overview of the current landscape of VR clinical training products and will serve as a valid reference for this analysis.

The full list of organisations and specific products is found in Table 7, and more information and references available in Appendix D.

Table 7 - List of commercial XR organisations and products

Organisation	**Relevant Product(s)**
E Learning Design Center (2022)	MedVR Education
GigXR (2022)	HoloScenarios
Health Scholars (2022)	Hospital VR, EMS VR
i3 Simulations (2022)	Resuscitation VR
inciteVR (2020)	Medical Assisting Clinical Suite
Lucid Reality Labs (2020)	Medical VR Intubation VR Hysteroscopy Multiplayer
Lumeto (2021)	Involve XR
Oxford Medical Simulation (2022)	OMS Nursing, OMS Medical
PCS (2022)	PCS Spark
SimX (2021)	SimX EMS
UbiSim (2022)	UbiSim VR
VRpatients (2022)	VRpatients for Nursing

5.3.1 Clinical Actions

To gain a comprehensive understanding of the *interaction designs* of these clinical training products, it is essential to first identify the expected clinical skills in a clinical training scenario. Al-Elq (2007) provides a list of such skills, defined as:

1. History Taking
2. Physical Examination
3. Clinical Investigation
4. Diagnostic Reasoning
5. Procedural Perfection
6. Effective Communication
7. Teamwork
8. Professionalism

The identified clinical skills can be further granulated into specific actions (Table 8), informed through discussions with clinicians at Children's Hospital Los Angeles and through observations of emergency medicine shifts. These actions provide a more in-depth understanding of the specific actions involved in these clinical skills, allowing for a more comprehensive analysis of the *interaction designs* in the commercial products.

Table 8 - List of clinical actions

Skill	**Actions**
1. History Taking	a. *Asking Patient Questions* b. *Asking Patient Companion Questions* c. *Asking Healthcare Personnel Questions*
2. Physical Examination	a. *Examining Patient* b. *Manipulating Patient Physically* c. *Note Taking*
3. Clinical Investigation	a. *Ordering Medical Tests* b. *Ordering Medical Scans* c. *Performing Medical Tests*
4. Diagnostic Reasoning	a. *Reading Patient Vitals* b. *Reading Patient Information* c. *Developing an Action Plan*
5. Procedural Perfection	a. *Administering Medication* b. *Administering Airway Management* c. *Administering Medical Equipment*
6. Effective Communication	a. *Communicating with Patient* b. *Communicating with Patient Companion* c. *Communicating with Hospital Departments*
7. Teamwork	a. *Closed-Loop Communication* b. *Giving Instructions* c. *Role Assignment*
8. Professionalism	a. *Following Protocol* b. *Empathetic Communication* c. *Risk Analysis*

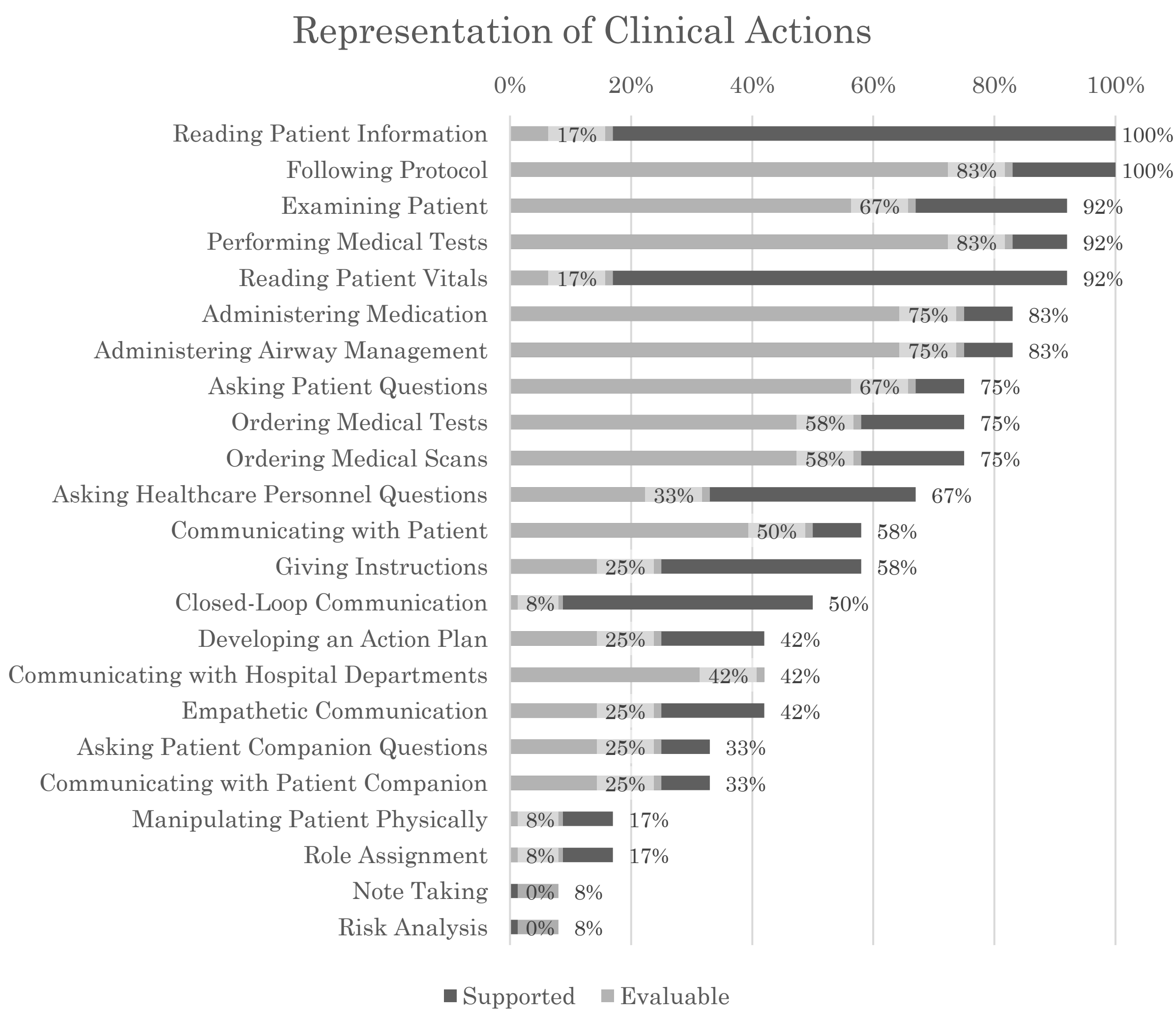
In this study, the 12 identified commercial VR clinical training products will be analysed based on the list of actions to evaluate what is supported in the simulations and what *interaction design* is utilised for each, as well as any evaluation frameworks. An interaction matrix will be constructed to illustrate the interactions supported by each product, and to identify common design patterns across the products.

To analyse the *interaction designs* of these products, a range of sources will be utilized, including product videos, public demonstrations, application trials, as listed in Appendix D. While it should be noted that these products may be subject to change, and comprehensive walkthroughs of the software may not be available, this evaluation will nonetheless provide valuable insights and highlight areas for potential improvement. Any limitations or gaps in the evaluation will be addressed in

5.3.2 Action Results

The list of commercial VR clinical training products was referenced against the list of clinical skills actions from Table 8. A summary of these results, ordered by support, is presented in Table 9, in full in Appendix E, and an interaction matrix can be found in Appendix F.

Table 9 - Clinical actions results



Here ‘Supported’ means that clinical actions are able to be triggered by the user, and ‘Evaluable’ means that clinical actions are scored, or feedback given on as part of simulation debriefing (for parity with manikin-based simulation). This distinction is made because the ability to assess a clinical skill is part of corresponding manikin-based simulations. Hence, classifying an action as 'Evaluable' indicates that it's not just supported, but a fundamental part of the complete training process, which may include feedback or scoring mechanisms, forming the feedback part of the interaction loop between user and system.

The most supported clinical skills action was *Following Protocol* and *Reading Patient Information*, which were represented in some capacity by all 12 simulations. *Following Protocol* was evaluated through the ability to read protocol documentation within the simulation, feedback from other healthcare professionals, and procedure checklists. *Reading Patient Information* was represented as either virtual props or UI menus from which patient background, histories, etc. could be accessed and read.

All but two of the simulations provided a means to evaluate *Following Protocol* directly, while the remaining two simulations were instructor-driven and required live supervision for evaluation. Conversely, *Reading Patient Information* was only evaluable in two simulations, along with *Reading Patient Vitals* (supported in 11, evaluable in 2), suggesting a lack of well-defined *interaction design* for capturing reading patient vitals and information as part of clinical practice. These actions are usually passive performances and were often represented by a simple UI menu or physical prop from which the trainee could read information.

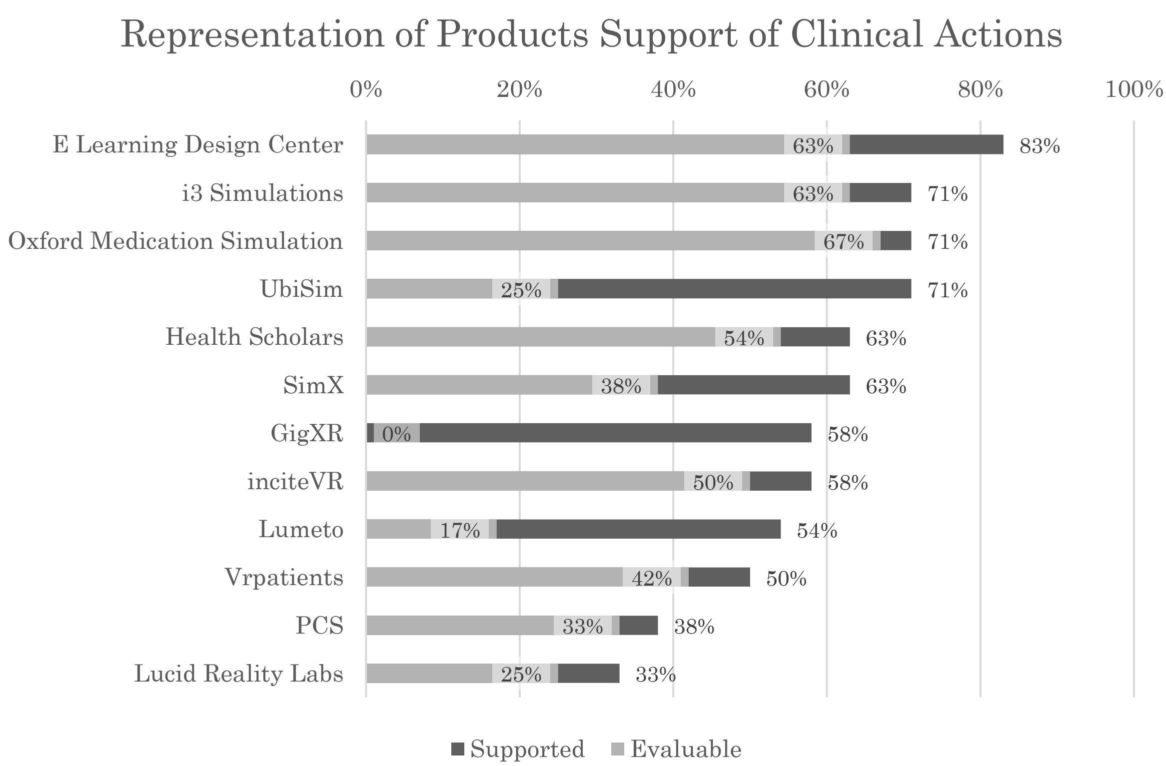
Examining Patient and *Performing Medical Tests* were also highly supported with 11 simulations supporting each. These actions were evaluable in most simulations, likely as they are "active" tasks, and were commonly represented through either virtual physical props or pointer-based menu selection. The virtual physical props were grabbed using the user controller and then activated at patient hotspots for performing examinations and tests in a pseudo-natural manner. Meanwhile, pointer-based menu selection allowed the user to select an examination or test from a list of options, which would then be performed automatically, and the results communicated to the user.

The two clinical skills actions with the least support were *Risk Analysis* and *Note Taking*, with only one simulation each representing these actions and neither being evaluable. In the Lumeto simulations, *Risk Analysis* was presented as a prompt for multi-user voice discussion, while *Note Taking* was observed in the inciteVR simulations, where a virtual whiteboard was utilized for writing purposes with the use of a virtual marker pen.

Other clinical skills actions that were found to be underrepresented in the simulation products analysed were *Manipulating Patient Physically*, and *Role Assignment*. These skills were supported by only two simulations each, and only one of them was evaluable. Moreover, the clinical skills actions of A*sking Patient Companion Questions* and *Communicating with Patient Companion* were represented in only four simulations each, yet they were evaluable in a significant proportion of these instances.

Table 10 shows the breakdown per organisations' products, ordered by support, and full data can be found in Appendix G.

Table 10 - Organisation action results



5.3.3 Interaction Designs

In the analysis of the commercial VR clinical training products, three common selection patterns (Jerald, 2016, sec. 28.1) were identified. These include Pointing Pattern, Hand Selection Pattern, and Non-Spatial Pattern, and were often found combined within simulations (known as Multimodal Pattern) depending on the interaction context:

- Pointing Pattern involves the use of a virtual laser emitted from the player's controller to point at and select virtual objects, menus, and UI elements. This interaction scheme may offer a direct and intuitive way for the user to interact with the virtual environment.

- Hand Selection Pattern, on the other hand, requires the user to physically grab and place objects and items to perform tasks. This interaction scheme is represented by a virtual hand on the user's controller and may allow for a more immersive and hands-on experience.
- Finally, Non-Spatial Pattern through voice control involves the use of the headset microphone by the user to give instructions, ask questions, and query the simulation. This interaction scheme is supported by AI-based Natural Language Processing (NLP) and may provide a hands-free option for interacting with the virtual environment.

Image redacted for
copyright / licensing reasons

Figure 4 - Pointing Pattern (i3 Simulations, used with permission) and Hand Selection Pattern (SimX)

Where Non-Spatial voice control and either Pointing or Hand Selection was found, it was typically as Concurrent Input Modalities (Jerald, 2016, sec. 26.6) where "users [can] different commands simultaneously", for example pointing to a medication to select whilst verbally requesting information from the patient. This has similarity to real-world simulation practice and will be explored further in Chapter 6.1.

It is also worth exploring object affordances in the VR simulations. Affordances, as outlined by Norman (2013, p. 10), are the relationships between the properties of an object and how a user understand the object's intended and available usages. This limitation of information attributed to objects in VR simulations has a significant impact on object affordances and the user's ability to understand the potential use of an object.

Without sufficient information - for example, feel the weight of an object, how it fits within their hand, how it can be manipulated physically - the user is unable to accurately understand the potential use of an object in the virtual environment, leading to the *gulf* *of execution*: difficulties in their ability to manipulate and interact with the objects. For this, haptics technology is a potential research area of interest for this problem, as explored in 7.1 Future Work.

Some knowledge of clinical objects to be physically used can be expected of trainee clinicians, but as seen in the literature review and user study, there is still the *gulf of execution* in which users struggle to understand how to use an object within the VR simulation. To address this challenge, some commercial VR products have implemented design patterns aimed at simplifying object usage:

- One example is automated object usage, which is seen with pointer-based interaction schemes. In this mode, the system automatically uses objects selected in the environment, eliminating the need for the user to learn how to interact with the tools.
- Another design pattern is snap-to-place tool usage, which is seen with hand-grab selection. This mode allows users to pick up items and move them to a location of expected use, where they automatically "snap" into place and are activated.
- A third design pattern is visual-guide tool usage, which is also seen with hand-grab selection. In this mode, visual "hotspots" are displayed when the user picks up an item, guiding them to the available locations of expected use.

Images redacted for
copyright / licensing reasons

Figure 5 - Visual-guide examples (SimX, inciteVR, UbiSim)

While these design patterns simplify tool usage, they also limit realism and may impact negatively on knowledge transfer. For example, they may skip important procedural steps, such as patient positioning and preparation. As a result, it may be important to consider unguided affordances in VR simulations to ensure that procedural knowledge and skills are effectively transferred from the simulation to actual clinical practice. This aspect will be explored in further detail in Chapter 7.1 of this thesis.

5.4 Evaluation

To partly answer Research Question 3 ('What are potential *interaction design* alternatives that can minimise usability issues?'): this market analysis found further gaps in simulation affordances that could impact usability issues due to the 'gulf of execution' but has also found common *interaction design* patterns that merit further investigation.

It was observed that despite being text-heavy actions that utilise virtual documentation *Risk Analysis* and *Note Taking* were underrepresented in most computer simulations, which is in contrast with *Following Protocol* and *Reading Patient Information*, which are also text-heavy but well supported. The lack of representation in simulation is speculated to be due to the difficulties in designing interactive and immersive *interaction designs* for these actions within a virtual reality (VR) environment. As outlined by Bowers et al. (2021), usability issues with the common interaction metaphor of 'air-drawing' for writing and drawing in VR have been identified, along with potential workarounds. With advancements in AI Natural Language Processing, new affordances for such documentation skills could be possible; this is explored further in 7.1 Future Work.

It is surprising to find that *Manipulating Patient Physically* was underrepresented in the VR clinical training simulations, as physically interacting with a patient is a common occurrence in clinical practice, particularly for diagnostic and therapeutic purposes. The user study conducted in Chapter 4 revealed that many participants expected to be able to move the patient into a recumbent position, for instance, despite the lack of support for this action in the simulations.

It is speculated that the underrepresentation of this action in VR clinical training products might be attributed to technical software limitations. The requirement of realistically rendering a patient with full physics, in a manner that will run smoothly on VR headsets, can be challenging due to performance issues. Nevertheless, it should be noted that advancements have been made in the field of VR surgical simulation with regards to "soft body" manipulation and deformation, which suggests that this underrepresentation might be solvable. Further investigation may be necessary to fully understand the limitations and potential solutions for incorporating this action in VR clinical training.

It is noteworthy that the clinical actions related to patient companions, such as *Asking Patient Companion Questions* and *Communicating with Patient Companion*, were underrepresented in the simulations analysed. Despite the significance of companions in facilitating communication and diagnosis, particularly with patients who are under 18 years of age, those with developmental disabilities, and elderly patients, these actions were only supported by four simulations each, although evaluable in most of these examples. Research has shown that companions play a crucial role in the hospital care process, especially during emergency admission and acute care (Bocchi *et al.*, 2007; Oliveira *et al.*, 2022), and that the average inpatient has at least one companion (Hung et al., 2020).

While there has been previous research into companion communication skills in virtual simulations (Sanders et al., 2021), virtual reality simulations have yet to fully explore this aspect. This could be because virtual reality simulations typically emulate manikin-based simulations, which do not typically feature simulated patient companions. However, the concept of "hybrid" simulations, which incorporate both use manikins as well as standardized patient elements, is gaining traction (Peterson, Porter and Calhoun, 2020), and could impact on virtual reality simulation design as standards change.

It is worth noting a split between multiuser and single-user simulations was observed in the market analysis. Four simulations only facilitated single-user simulations, whereas most simulations (8) enabled multiuser simulations to some extent. However, this division was not found to have a substantial impact on the support for most clinical actions. There was a 38% higher support for the clinical skills actions of *Ordering Medical Tests* and *Closed-Loop Communication with Healthcare Personnel* in multiuser simulations, which may be because these actions are better facilitated with the involvement of multiple real-life users. However, it should be noted that these differences cannot be considered statistically significant and therefore should be explored further as a potential trend.

In the following Chapter 6, the findings from this market analysis, alongside the literature review and user study results, will be discussed and synthesised into design recommendations for ongoing development and future work.

Chapter 6 – Discussion and Recommendations

In Chapter 6, we discuss the findings of this project and outline future work that is needed to continue improving and advancing the effectiveness of VR clinical simulations discussed in this study. This discussion and recommendations aim to complete the answer for Research Question 3: What are potential *interaction design* alternatives that can minimise usability issues?

In this research, it has been noted that the limitations of extra learning affected the knowledge transferability from virtual reality (VR) simulations. This was observed in two areas of the VR experience. Firstly, when users were learning new interaction schemes, such as pointer- or grab-based interactions, there was a limitation in their ability to effectively use these controls. Secondly, the possibility space of the VR simulation was also a factor, with novice users and older users struggling to fully understand the full extent of what was possible within the simulation, as seen within the Literature Review.

In this project, a comprehensive review of the current state-of-the-art clinical training simulators using virtual reality technology was conducted. The findings revealed a generally high level of user acceptance among the clinician community; however, several usability issues were identified, particularly among novice users and older clinicians. To further understand these issues, a user study was carried out at Children's Hospital Los Angeles, using the Resuscitation VR software, which is designed for training clinical skills in an emergency medicine scenario. The results of the user study highlighted similar outcomes as the literature review and shed light on the issue of the "gulf of execution" (Norman, 2013), which is a limiting factor for usability.

To identify areas for improvement and understand the impact of these usability issues on the current commercial VR clinical simulation market, a market analysis was performed on commercially available VR clinical simulation products. The analysis focused on the interaction schemes and limitations, especially when compared to actual clinical practice actions. The findings of the analysis revealed several areas for improvement, linked with findings from both the literature review and user study. These findings will be explored in the following chapter as a basis for future work.

These findings highlight the importance of effective learning design in virtual reality simulations, particularly in terms of the interactions users must undertake to effectively use the simulation. The limitations of extra learning on knowledge transferability also have implications for the development and implementation of VR simulations for clinical training purposes.

6.1 Human-Centred Design Recommendations

To answer Research Question 3 ('What are potential *interaction design* alternatives that can minimise usability issues?') and to meet the identified usability issues and limitations in VR clinical training simulations, several *interaction design* alternatives have been identified by a synthesis of the findings to RQ1 & 2, and from an examination of current VR design literature. A summary of these potential improvements can be found in Appendix H.

Firstly, extra learning limits knowledge transferability from the VR simulations. This was found with both learning new control schemes, be it pointer- or grab-based interactions, and with learning the *possibility space* of the VR simulation – particularly again with novice users and older clinicians, as suggested in the literature review findings.

Onboarding tutorialisation could be enhanced with 'gating' design, to introduce interface interactions to novice users as seen with novel game hardware (Dabic, Lund and Nova, 2010) and serious games (Shum *et al.*, 2023).

Exploring using feedforward – using visual cues to demonstrate the affordances and outcomes of using an item – as discussed by Muresan, McIntosh and Hornbæk (2023) for tutorialisation could also prove useful as clinical trainees are encountering many medical tools and objects in the virtual space. Explicit feedforward previews are suitable for tutorials, but subtle, 'hidden' feedforward previews may be explored for deployment within the clinical simulations themselves.

There is some discussion that users generally hold high expectations for interaction affordances within VR applications (Weber, Weibel and Mast, 2021; Rauschnabel *et al.*, 2022). Increasing the number of interaction options to meet Uncoded category user errors might be counterproductive however, potentially leading to more user errors and clutter. Two design strategies arise: adding virtual objects/interfaces, possibly causing user errors and visual clutter, or replacing virtual objects/interfaces, which could bias user selections. Some user actions, revealed during the Think-Aloud sessions, were uncoded as they were also not expected within manikin-based training, e.g., patient companion communication, patient repositioning, after-care requests, etc. Responding to such user expectations could enhance the VR experience.

Controller Issues could be improved with the investigation of hand tracking as a potential solution to controller issues. While such interfaces were once limited by high cost hardware (Wozniak *et al.*, 2016; Strazdins *et al.*, 2017) or limited computer vision algorithms (Schlattman and Klein, 2007; Wang and Popović, 2009; Pan *et al.*, 2010), hand-tracking functionality is now accessible as standard for the Meta Quest headset (Meta, 2020).

Research has shown that hand tracking has higher usability than virtual reality controllers for certain tasks such as object manipulation and typing (Voigt-Antons *et al.*, 2020), and this should be validated for use with immersive clinical training simulations specifically. Although hand-tracking can mitigate some issues, it doesn't solve Selection errors. Proximity Selection or Direct Hand Selection (Jerald, 2016, sec. 28.1) might be alternatives. If 'voice control' can reduce controller-based errors, Hand Selection can further reduce Selection Errors by allowing users to 'grab' and 'touch' objects as in real-life.

This could also be mitigated with the use of haptic devices, which have already found success in being incorporated into VR surgical simulations (Rangarajan, Davis and Pucher, 2020). To improve hand-tracked interaction, design elements recommended by Masurovsky et al. (2020) include smart object colouring, open-hand gestures for grabbing objects, and audio feedback to enhance user interactions, and the avatar hands themselves should include 1-to-1 finger movements to increase embodiment, with a high-realism hand model only used with high interaction fidelity to negative influence on experience caused by unmet expectations (Zhang *et al.*, 2023).

To address the State Misidentification issues seen in the user study, visual interfaces could be designed that allow trainees to document the steps and medications they have already taken. This documentation should be easily accessible and prominently displayed within the VR simulation. Additionally, incorporating more options for information requests from other staff in the clinical scenario can help to address any uncertainties that trainees may have. Currently, information exchange is actioned through in-simulation UI menus, for which Pfeuffer *et al.*, (2020) recommend integrating gaze input with handheld menus, as gaze-based interaction performs similarly to pointer-based selection but requires less physical effort, offering a more efficient and user-friendly interaction approach. Additionally, Yildirim (2022) finds voice cues, rather than text cues, to increase knowledge retention by users, and as such should be considered when design information exchange interfaces.

This study underscores the challenge of replicating Closed-Loop Communication, common in other medical simulations, within VR (Cordar *et al.*, 2017; Balint, 2019). A potential solution could be a 'voice control' interface using Natural Language Processing (Jerald, 2016, sec. 26.4.2), which has already been adopted in some VR medical applications (Ingrassia *et al.*, 2020; Katz *et al.*, 2020).

Where voice control is available, these requests could be in the form of Natural Language Processing (NLP) interpolation, or otherwise visual conversational interfaces that provide trainees with the ability to request necessary information they need to make informed decisions. This functionality enables an expansion of available options without triggering the same *interaction design* errors. It also affords user feedback on uncoded actions, like those in Rule-Based Errors. Furthermore, it could reduce State Misidentify errors as users could use 'voice control' to request clarification on their current simulation understanding.

Where communicating with virtual agents such as members of the clinical team, interaction design should incorporate the feedback from these agents, for which (Wei, Jin and Fan, 2022) suggest symbolic (i.e., nodding) and natural (i.e., pointing) gestures to express agent intentions and enhance communication and joint attention. These visual cues minimize verbal communication needed, increasing efficiency, and decreasing feedback loop time.

The implementation of voice control via speech recognition (Jerald, 2016, sec. 26.4.2) using Natural Language Processing across the board should additionally be considered as it is an intuitive and natural way of interaction in a clinical environment and aligns with expectations brought over from manikin-based training. The use of voice control would allow for more natural interaction via Concurrent Input Modalities (Jerald, 2016, sec. 28.5.3) with the virtual environment and could reduce the 'gulf of execution' and improve usability, which has been explored in recent research in clinical VR simulations (Katz *et al.*, 2020; Sapkaroski, Mundy and Dimmock, 2022).

By combining voice commands with other modal information, such as gesture and gaze, user intention understanding can be improved, leading to more realistic and immersive VR experiences (LI et al., 2019). 'Physical' actions on virtual objects are still required within clinical simulations, and as such frequent Controller Issues may remain. This study found those mostly arise from the controller itself (Interface Confusion, Accidental Press) and the in-simulation control (Selection Error). A multimodal interface with voice and hand-tracking should be explored. Tao et al. (2021) emphasize the importance of maintaining a balance between intuitive actions (e.g., hand-tracking, voice commands) and simple, reliable commands (e.g., button presses) for multimodal inputs.

AI-based NLP may also be implemented for the purposes of intuitive documentation clinical actions, such as *Risk Analysis* and *Note Taking*, for providing intuitive feedback for when clinical actions and medications requested are uncoded as seen in Chapter 4 - User Study Results and Analysis, and for incorporating patient companions into all scenarios for communication skills training to provide a more realistic and effective training experience, as seen in Chapter 5.

Despite noted usage of NLP voice control in the evaluated market products, some issues such as limited conversational and language comprehension (Ng *et al.*, 2022) limit the possibility space of NLP usage at present, so more research and development may be necessary to fully utilise this functionality.

Additionally, within the collected VR clinical applications there were not found any compound selection patterns (Jerald, 2016, sec. 28.5) such as Point Hand Pattern. This combines advantages of Pointing Pattern and Hand Grab Pattern, so that "objects are first selected via pointing and then manipulated as if held in the hand". This may be appropriate for VR simulations lacking hand tracking or voice control functionality as it would allow for quicker interaction cycles.

Kang, Shin and Ponto (2020) advocate for a 'worlds-in-miniature' *interaction design* that leverage visual affordance and physics to facilitate accurate object manipulations, which would pair well with Point Hand Pattern. This design appears to be novel within VR *interaction design* research (Widjojo, Chinthammit and Engelke, 2017; Sivakumar *et al.*, 2021), so this recommendation should be explored as a low-effort alternative to improving usability.

Furthermore, the underrepresented skills of Risk Analysis and Note Taking should also be considered for addition to existing VR clinical training simulations. These skills are crucial in actual clinical practice and could easily be integrated into the existing products to bridge the gap between simulation and reality. A stopgap may be *interaction design* for handwriting in VR. Research has shown some VR handwriting patterns to be more engaging than virtual keyboards, but with worse overall input performance (Elmgren, 2017). This is confirmed by (Hellum, Kersten-Oertel and Xiao, 2023), who also found that controller-based interfaces were preferable to gaze-based *interaction design*. Apart from NLP voice control as noted above, additional research may be required to identify the best *interaction design* for handwriting in VR space to fully integrate these clinical actions in an intuitive manner where NLP is not suitable. Tu *et al.*, (2019) suggest the use of crossing selection techniques for interfaces in VR, aided by visual or audio guides, as an alternative to pointing, which could improve the usability issues seen with VR text inputs. Similarly, multimodal (auditory and haptic) feedback has been shown to improve VR text input performance as per Yildirim (2022).

To further enhance the efficacy of VR clinical training, the potential for eye tracking technology should be investigated. Eye tracking has become available in commercial VR hardware with the release of hardware like the HTC Vive Pro Eye (HTC Corporation, 2019) and Meta Quest Pro (Meta, 2022). It could serve as a valuable tool for evaluating body language for empathetic communication, and observation skills for diagnosis training. Choi and Nam (2023) have found a Circle Marker sphere-cast visualisation as the most effective for displaying user gaze with eye tracking, which displays a subtle visual guide for users to understand what they will be interacting with. Gaze-based interactions would also benefit from multimodal feedback as per Adhanom, MacNeilage and Folmer's (2023) review.

Overall, these improvements have the potential to enhance the realism and effectiveness of VR clinical training simulations, reducing the usability issues and limitations identified in the user study and market analysis. Görlich, Akincir, and Meixner (2022) highlight the need for a comprehensive collection of *interaction design* for VR simulations. Exploring and validating these designs would facilitate the dissemination of proven improvements and support the development of future clinical VR training with enhanced usability and user experiences.

In summary, key recommendations include the incorporation of intuitive visual communication for both object affordances, and also that allow trainees to document their actions and medications in the VR simulation. A greater emphasis should be placed on incorporating voice cues for information exchange interfaces, as they have been shown to increase knowledge retention. The use of AI-based Natural Language Processing for voice control and conversational interfaces should be explored to enhance trainee communication with virtual agents. Implementation of voice control across the VR simulations would provide a more natural interaction and improve usability, thereby reducing the 'gulf of execution'.

Where voice control isn't possible, potential solutions could be the exploration of hand tracking and the integration of haptic devices. In addition, for VR simulations lacking hand tracking or voice control functionality, the application of the Point Hand Pattern could be useful to improve interaction cycles. The addition of underrepresented skills such as Risk Analysis and Note Taking in VR clinical training simulations should also be considered.

The exploration of eye-tracking technology could also offer a viable tool for enhancing empathy and observational skills in diagnosis training. Lastly, more research is needed to ensure the successful integration of these recommendations, addressing limitations such as language understanding in NLP and the learning curve of new control schemes, and to further improve the usability and realism of VR clinical training simulations.

Ultimately*, interaction design* should focus on enabling seamless translation of skills from virtual environments to real-world practice. By incorporating these recommendations, designers and researchers can contribute to creating more immersive, interactive, and effective VR experiences in various domains.

Further research and development in these areas are necessary to fully realize the potential of VR in clinical training and future work to achieve this is explored in 7.1 Future Work.

Chapter 7 - Conclusion

This thesis aimed to evaluate current *interaction design* and such impacts on usability in VR clinical training applications through a literature review, a user case study, and a market analysis of commercially available VR applications in the domain.

This project sought to address three key research questions:

- RQ 1: What is the current state and limitations of virtual reality clinical skills training applications in healthcare?
- RQ 2: What are the barriers to usability for virtual reality simulations in clinical training?
- RQ 3: What are potential *interaction design* alternatives that can minimise usability issues?

The literature review, addressing RQ 1, reveals that despite VR's potential in healthcare training, challenges exist in control systems and hardware. These issues interrupt immersion, reducing training quality (Schild, Elsenbast and Carbonell, 2021). To leverage VR as a feasible simulation tool, it is imperative to enhance the intuitiveness of controls for novice users (Chang et al., 2019). Unfamiliarity with VR may also lead to lower test and usability scores, dampened enthusiasm, and dissatisfaction, and there is a learning curve with the new technology that impacts knowledge transferability to the VR simulator. The review suggests that utilizing hand-tracking systems and including verbal communication skills in simulation training can improve usability.

Answering RQ 2, the user study underscores that the primary usability barriers in VR simulations for clinical training arise from limited information exchange, rather than interaction control issues. It also highlights the persistent 'gulf of execution' (Norman, 2013), denoting the range of permissible user interactions, or 'possibility space' (Bogost, 2008). The study reported medium workload and satisfactory usability using the NASA-Task Load Index and System Usability Scale. However, user errors correlating with usability scores were found, hinting at VR's limitations in fully replicating manikin-based training's decision-making affordances. Despite the current interaction design being suitable as an adjunct to manikin-based training, other user errors were identified. These errors, which are tied to Norman’s (2013) concept of the "gulf of execution", were the only ones found to have a significant correlation with usability scores. This emphasizes the limitation in fully emulating the decision-making affordances provided by manikin-based training with the current VR *interaction design*.

Addressing RQ 3, a market analysis identified gaps in simulation affordances impacting usability due to the 'gulf of execution', alongside common *interaction design* patterns warranting further study and several interaction design alternatives were recognized. The analysis presents an interaction matrix of each product evaluated for its interaction design. The results revealed the most and least supported clinical skills actions in VR medical simulation, and the underrepresented actions in the analysed products. The underrepresented clinical skills in the VR medical simulation, such as manipulating patients physically and risk analysis, are speculated to be due to technical software limitations and difficulties in designing immersive interaction designs.

By implementing the recommendations made in this research, future work for VR high-fidelity simulations can improve usability, increase adoption of VR technology, and meet the demands of high-fidelity training needs identified by the Health Education England National Strategic Vision for simulation and immersive technologies in health and care (Health Education England, 2020). Overall, this study highlights the importance of evaluating VR in clinical skills training and provides insights for improving the design of VR training applications.

This thesis's findings can guide the creation of enhanced, user-centric VR training applications, with a goal is to bolster healthcare professionals' training and preparedness, thereby leading to improved patient outcomes. It is hoped that the outcomes and suggestions from this research will have broader impact on wider clinical VR simulation research and design, leading to future work that target usability issues identified in existing simulations.

Potential areas of exploration were targeted for a Human-Centred Design (Jerald, 2016) to reduce this identified 'gulf of execution'. These include implementing feedforward previews for object affordance conveyance, investigating hand tracking as a solution to controller issues, voice control via NLP for more natural interaction, the integration of intuitive visual interfaces for action documentation, exploring compound selection patterns such as Point Hand Pattern, and investigating eye-tracking technology. Additionally, interaction design for handwriting in VR may be necessary, and incorporating underrepresented skills such as Risk Analysis and Note Taking. These improvements may bridge the 'execution gap', providing a more effective training experience.

The concluding section is a list of potential future work that has been identified as necessary to achieve the recommendations outlined in this thesis, as well as common threads for additional investigation.

7.1 Future Work

There are some potential areas of usability improvement that could also be afforded by additional research studies.

Within the remit of *interaction design* is multiuser collaboration, as seen frequently in the market analysis in Chapter 5. Future studies could investigate the effect of avatar – both user and non-user characters - design and representation on multiplayer interaction intuition and effectiveness in VR clinical training simulations. Previous research has indicated that VR avatar appearance can influence attitudes and motivation (Hudson and Hurter, 2016), but little is known about the impact of avatar design on clinical training scenarios. Further studies could explore how factors such as avatar realism, gender, and ethnicity affect user perception of clinical scenarios, communication, and collaboration.

Additionally, studies could investigate how the design of avatars can enhance or hinder the acquisition of clinical skills and knowledge, and how the design of virtual environments and avatars can facilitate interprofessional collaboration and communication in clinical settings. Such studies could provide insights into how avatar design can be optimized to improve the effectiveness of multiplayer interactions in VR clinical training simulations.

To fully realize the potential of voice control via natural language processing (NLP) in clinical VR simulations, further research is needed to address current limitations. The use of AI-based NLP for intuitive documentation clinical actions and for incorporating patient companions into scenarios requires the development of comprehensive and dynamic conversation models that can accurately understand and respond to natural language commands. Additionally, research is needed to identify the best ways to integrate NLP with other potential modalities, such as hand tracking, to create a seamless and natural interaction experience.

Further research could investigate the effectiveness of unguided affordances in VR simulations for improving knowledge transferability of clinical object usage. While automated, snap-to-place, and visual-guide design patterns simplify tool usage, they may also skip important procedural steps and limit realism. Evaluating the impact of unguided affordances on transferability can provide insight into how to effectively transfer procedural knowledge and skills from VR simulations to actual clinical practice. Additionally, future studies could examine how to balance simplicity and realism in VR training simulations to optimize skill transfer.

Utilising haptic devices to mitigate the 'gap of execution' in VR clinical actions is another avenue of research. As the resource costs of haptic devices continue to decrease and the fidelity of haptics increases, investigating their use in VR clinical simulations for interacting with patients and objects will be crucial for widespread adoption. Haptic devices have already been successfully integrated into surgical simulations (Rangarajan, Davis and Pucher, 2020), and it is important to evaluate their effectiveness in clinical training scenarios. Interaction design research will be necessary to effectively incorporate haptic devices into clinical training in VR, as well as to determine the optimal level of haptic feedback necessary for effective training outcomes.

In regards to eye tracking, while its effectiveness has been explored in surgical VR training (Parham *et al.*, 2019; Mikhailenko, Maksimenko and Kurushkin, 2022), there has been little research conducted on its potential for improving clinical skills in VR. A promising future study would be to incorporate eye-tracking design into the user study application discussed in Chapter 3 - with a particular focus on addressing clinical skills evaluation gaps highlighted in Chapter 5 - and performing a similar usability study.

Future work could also include a meta-analysis of the evaluation methods employed in virtual reality (VR) research. While previous studies have evaluated the effectiveness of VR simulations for clinical training, as seen in Chapter 2, there is a lack of consistency in the evaluation methods used across studies, which makes it difficult to compare results and draw meaningful conclusions. Therefore, a meta-analysis of existing literature on VR evaluation methods can provide insights into the effectiveness of different evaluation methods and identify areas for improvement. This meta-analysis should focus on identifying the most effective evaluation methods for assessing clinical training outcomes in VR simulations, particularly in comparison to evaluation methods for traditional training simulations and provide recommendations for standardization of evaluation methods to ensure comparability and reproducibility of results.

In Chapter 5 - Market Analysis a split was seen in the products of multiuser and single-user VR simulations. While there has been some research (Wei, Jin and Fan, 2022) towards multiuser versus single-user VR simulation design in training, more investigation is necessary to fully understand the potential benefits and drawbacks of each approach.

Multiuser VR simulations offer the potential for collaborative learning and communication skills training but may also introduce additional usability challenges related to coordination and interaction between users. Single-user VR simulations may be more straightforward in terms of usability but may limit opportunities for communication skills training and other collaborative activities. A comparative study that measures usability and learning outcomes between the two approaches would be beneficial in informing design decisions for VR clinical training simulations.

Additionally, as identified in Chapter 2, it will be essential to evaluate the effectiveness of different VR onboarding methods, such as hands-on tutorialisation, external videos, and written guides, in achieving knowledge transferability. This could be done by conducting controlled studies with participants randomly assigned to different training methods and then measuring their performance in VR clinical simulations. The results could then be analysed to identify the most effective training method or combination of methods. This evaluation will provide insight into the most efficient and effective way to train clinicians to use VR simulations and improve knowledge transferability, which is crucial for successful implementation of these technologies in clinical practice.

Finally, to ensure that VR clinical training simulations are accessible and usable for all learners, including those with disabilities, it will be important to conduct usability testing with a wider demographic. This would involve recruiting participants with a range of physical, sensory, and cognitive impairments, and evaluating how well they are able to interact with the simulations. Specific accessibility features, such as text-to-speech, haptic feedback, or alternative control schemes, could be developed and tested to improve the usability of the simulations for learners with disabilities. The results of this testing would help to identify any barriers to accessibility and provide insights into how these can be overcome, ensuring that the simulations are inclusive and accessible to all learners.

References

Abulfaraj, M.M. *et al.* (2021) 'Virtual Reality vs. High-Fidelity Mannequin-Based Simulation: A Pilot Randomized Trial Evaluating Learner Performance', *Cureus*, 13(8), p. e17091. Available at: https://doi.org/10.7759/cureus.17091.

Adams, H. *et al.* (2019) 'Development and evaluation of an immersive virtual reality system for medical imaging of the ear', in B. Fei and C.A. Linte (eds) *Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling*. SPIE, p. 36. Available at: https://doi.org/10.1117/12.2506178.

Adhanom, I.B., MacNeilage, P. and Folmer, E. (2023) 'Eye Tracking in Virtual Reality: a Broad Review of Applications and Challenges', *Virtual Reality*, 27(2), pp. 1481–1505. Available at: https://doi.org/10.1007/s10055-022-00738-z.

Adhikari, R. *et al.* (2021) 'A mixed-methods feasibility study to assess the acceptability and applicability of immersive virtual reality sepsis game as an adjunct to nursing education', *Nurse Education Today*, 103, p. 104944. Available at: https://doi.org/10.1016/j.nedt.2021.104944.

Al-Elq, A.H. (2007) 'MEDICINE AND CLINICAL SKILLS LABORATORIES', *Journal of Family and Community Medicine*, 14(2), pp. 59–63. Available at: https://journals.lww.com/jfcm/Fulltext/2007/14020/MEDICINE_AND_CLINICAL_SKILLS_LABORATORIES.4.aspx.

Al-Ghareeb, A.Z. and Cooper, S.J. (2016) 'Barriers and enablers to the use of high-fidelity patient simulation manikins in nurse education: an integrative review', *Nurse Education Today*, 36, pp. 281–286. Available at: https://doi.org/10.1016/j.nedt.2015.08.005.

Alger, M. (2015) *Visual Design Methods for Virtual Reality*. Available at: http://aperturesciencellc.com/vr/VisualDesignMethodsforVR_MikeAlger.pdf.

Almousa, O. *et al.* (2019) 'Virtual Reality Simulation Technology for Cardiopulmonary Resuscitation Training: An Innovative Hybrid System With Haptic Feedback', *Simulation & Gaming*, 50(1), pp. 6–22. Available at: https://doi.org/10.1177/1046878118820905.

Balint, B.N. (2019) 'Designing VR for teamwork: The influence of HMD VR communication capabilities on teamwork competencies', *26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings*, pp. 1365–1366. Available at: https://doi.org/10.1109/VR.2019.8798147.

Baniasadi, T., Ayyoubzadeh, S.M. and Mohammadzadeh, N. (2020) 'Challenges and Practical Considerations in Applying Virtual Reality in Medical Education and Treatment', *Oman Medical Journal*, 35(3), pp. e125–e125. Available at: https://doi.org/10.5001/omj.2020.43.

Barteit, S. *et al.* (2021) 'Augmented, Mixed, and Virtual Reality-Based Head-Mounted Devices for Medical Education: Systematic Review', *JMIR Serious Games*, 9(3), p. e29080. Available at: https://doi.org/10.2196/29080.

Berg, H. and Steinsbekk, A. (2020) 'Is individual practice in an immersive and interactive virtual reality application non-inferior to practicing with traditional equipment in learning systematic clinical observation? A randomized controlled trial', *BMC Medical Education*, 20(1), p. 123. Available at: https://doi.org/10.1186/s12909-020-02030-7.

Blueweave Consulting (2022) *Virtual Reality in Healthcare Market - Global Size, Share, Trend Analysis, Opportunity and Forecast Report, 2018–2028*. Available at: https://www.blueweaveconsulting.com/report/virtual-reality-in-healthcare-market.

Bocchi, S.C.M. *et al.* (2007) 'Family visitors and companions of hospitalized elderly and adults: analysis of the experience from the perspective of the nursing working process', *Revista Latino-Americana de Enfermagem*, 15(2), pp. 304–310. Available at: https://doi.org/10.1590/S0104-11692007000200017.

Bogost, I. (2008) 'The rhetoric of video games', *The Ecology of Games: Connecting Youth, Games, and Learning*, pp. 117–139. Available at: https://doi.org/10.1162/dmal.9780262693646.117.

Boletsis, C. and Cedergren, J.E. (2019) 'VR Locomotion in the New Era of Virtual Reality: An Empirical Comparison of Prevalent Techniques', *Advances in Human-Computer Interaction*, 2019. Available at: https://doi.org/10.1155/2019/7420781.

Boros, M. *et al.* (2022) 'Application of VR Technology to the Training of Paramedics', *Applied Sciences*, 12(3), p. 1172. Available at: https://doi.org/10.3390/app12031172.

Bowers, B. *et al.* (2021) 'Comparing Virtual Constraints and a Physical Stylus for Planar Writing and Drawing in Virtual Reality', in *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, pp. 220–225. Available at: https://doi.org/10.1109/VRW52623.2021.00048.

Bracq, M.-S., Michinov, E. and Jannin, P. (2019) 'Virtual Reality Simulation in Nontechnical Skills Training for Healthcare Professionals', *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, 14(3), pp. 188–194. Available at: https://doi.org/10.1097/SIH.0000000000000347.

Brooke, J. (1996) 'SUS - A quick and dirty usability scale', *Usability evaluation in industry*, 189(194), pp. 4–7. Available at: https://doi.org/10.1002/hbm.20701.

Brooke, J. (2013) 'SUS: A Retrospective', *Journal of Usability Studies*, 8(2), pp. 29–40. Available at: http://www.usabilityprofessionals.org/upa_publications/jus/2013february/brooke1.html%5Cnhttp://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html.

Bunz, U., Seibert, J. and Hendrickse, J. (2021) 'From TAM to AVRTS: development and validation of the attitudes toward Virtual Reality Technology Scale', *Virtual Reality*, 25(1), pp. 31–41. Available at: https://doi.org/10.1007/s10055-020-00437-7.

Calandra, D., Lamberti, F. and Migliorini, M. (2019) 'On the usability of consumer locomotion techniques in serious games: Comparing arm swinging, treadmills and walk-in-place', *IEEE International Conference on Consumer Electronics - Berlin, ICCE-Berlin*, 2019-Septe, pp. 348–352. Available at: https://doi.org/10.1109/ICCE-Berlin47944.2019.8966165.

Caravella, E. (2019) *Teaching Gamefully: Proceduralizing the Classroom Through Possibility Space Pedagogy*. George Mason University.

Carruth, D.W. (2017) 'Virtual reality for education and workforce training', in *2017 15th International Conference on Emerging eLearning Technologies and Applications (ICETA)*. IEEE, pp. 1–6. Available at: https://doi.org/10.1109/ICETA.2017.8102472.

Chang, T.P. *et al.* (2019) 'Comparisons of Stress Physiology of Providers in Real-Life Resuscitations and Virtual Reality–Simulated Resuscitations', *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, 14(2), pp. 104–112. Available at: https://doi.org/10.1097/SIH.0000000000000356.

Chang, T.P. *et al.* (2020) 'Development and Considerations for Virtual Reality Simulations for Resuscitation Training and Stress Inoculation', *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, 16(6), pp. e219–e226. Available at: https://doi.org/10.1097/SIH.0000000000000521.

Choi, H. and Nam, S. (2023) 'The Effect of the Gaze Visualization Interface on the Learning Performance of Virtual Reality Medical Education', *2023 8th International Conference on Business and Industrial Research (ICBIR)*, (May), pp. 507–511. Available at: https://doi.org/10.1109/icbir57571.2023.10147557.

Choi, K.-S. (2022) 'Virtual reality simulation for learning wound dressing: Acceptance and usability', *Clinical Simulation in Nursing*, 68, pp. 49–57. Available at: https://doi.org/10.1016/j.ecns.2022.04.010.

Cohen, L. *et al.* (2018) *Augmented and Virtual Reality in Operations*.

Colonna, A.L. *et al.* (2022) 'Trauma bay virtual reality—A game changer for ATLS instruction and assessment', *Journal of Trauma and Acute Care Surgery*, 93(3), pp. 353–359. Available at: https://doi.org/10.1097/TA.0000000000003569.

Cooke, L., Dusenberry, L. and Robinson, J. (2020) 'Gaming Design Thinking: Wicked Problems, Sufficient Solutions, and the Possibility Space of Games', *Technical Communication Quarterly*, pp. 1–14. Available at: https://doi.org/10.1080/10572252.2020.1738555.

Cooper, J.B. and Taqueti, V.R. (2008) ‘A brief history of the development of mannequin simulators for clinical education and training’, *Postgraduate Medical Journal*, 84(997), pp. 563–570. Available at: https://doi.org/10.1136/qshc.2004.009886.

Cordar, A. *et al.* (2017) ‘Repeat after me: Using mixed reality humans to influence best communication practices’, in *2017 IEEE Virtual Reality (VR)*. IEEE, pp. 148–156. Available at: https://doi.org/10.1109/VR.2017.7892242.

Crosby, L.E. *et al.* (2019) ‘An Immersive Virtual Reality Curriculum for Pediatric Providers on Shared Decision Making for Hydroxyurea’, *Blood*, 134(Supplement_1), pp. 3402–3402. Available at: https://doi.org/10.1182/blood-2019-128661.

Dabic, S., Lund, K. and Nova, N. (2010) ‘Effects of expertise, mapping and tutorial format on user experience when gesturing with the Nintendo Wii’, *IHM’10 - 22nd Conference Francophone sur l’Interaction Homme-Machine*, pp. 121–124. Available at: https://doi.org/10.1145/1941007.1941026.

Datta, R., Upadhyay, K. and Jaideep, C. (2012) ‘Simulation and its role in medical education’, *Medical Journal Armed Forces India*, 68(2), pp. 167–172. Available at: https://doi.org/10.1016/S0377-1237(12)60040-9.

Davis, F.D. (1989) ‘Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology’, *MIS Quarterly*, 13(3), p. 319. Available at: https://doi.org/10.2307/249008.

Dumas, J. (2001) ‘Usability testing methods: Think-aloud protocols’, in *Design by people for people: Essays on usability*. Usability Professionals’ Association, pp. 119–130.

E Learning Design Center (2022) *XR in healthcare | XR Healthcare Training Solutions*. Available at: https://medvr.education.

EBSCOhost (2022) *Advanced Search: EBSCOhost*. Available at: https://web.s.ebscohost.com/ehost/search/advanced.

EcoMap Technologies Inc (2022) *Explore Organisations*. Available at: https://xrecomap.com/organizations (Accessed: 3 March 2022).

Elmgren, R. (2017) *Handwriting in VR as a Text Input Method*. KTH Royal Institute of Technology.

Feick, M. *et al.* (2020) ‘The Virtual Reality Questionnaire Toolkit’, in *Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology*. New York, NY, USA: ACM, pp. 68–69. Available at: https://doi.org/10.1145/3379350.3416188.

Ferguson, J. *et al.* (2020) ‘Implementing, embedding and sustaining simulation-based education: What helps, what hinders’, *Medical Education*, 54(10), pp. 915–924. Available at: https://doi.org/10.1111/medu.14182.

Fournier, J.P., Demeester, A. and Charlin, B. (2008) 'Script Concordance Tests: Guidelines for Construction', *BMC Medical Informatics and Decision Making*, 8(1), p. 18. Available at: https://doi.org/10.1186/1472-6947-8-18.

Freina, L. and Ott, M. (2015) 'A literature review on immersive virtual reality in education: state of the art and perspectives', *Conference proceedings of eLearning and Software for Education (eLSE)*, (01), pp. 133–141.

GigXR (2022) *GigXR - Homepage*. Available at: https://www.gigxr.com.

Goodhue, D.L. and Thompson, R.L. (1995) 'Task-Technology Fit and Individual Performance', *MIS Quarterly*, 19(2), p. 213. Available at: https://doi.org/10.2307/249689.

Gordon, C.L. *et al.* (2019) 'Affordance Compatibility Effect for Word Learning in Virtual Reality', *Cognitive Science*, 43(6). Available at: https://doi.org/10.1111/cogs.12742.

Haerling, K.A. (2018) 'Cost-Utility Analysis of Virtual and Mannequin-Based Simulation', *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, 13(1), pp. 33–40. Available at: https://doi.org/10.1097/SIH.0000000000000280.

Han, S.G. *et al.* (2021) 'Virtual reality-based neurological examination teaching tool (VRNET) versus standardized patient in teaching neurological examinations for the medical students: a randomized, single-blind study', *BMC Medical Education*, 21(1), p. 493. Available at: https://doi.org/10.1186/s12909-021-02920-4.

Harrington, C.M. *et al.* (2018) 'Development and evaluation of a trauma decision-making simulator in Oculus virtual reality', *The American Journal of Surgery*, 215(1), pp. 42–47. Available at: https://doi.org/10.1016/j.amjsurg.2017.02.011.

Hart, S.G. (2006) 'NASA-task load index (NASA-TLX); 20 years later', *Human Factors and Ergonomics Society Annual Meting*, pp. 904–908. Available at: https://doi.org/10.1037/e577632012-009.

Hart, S.G. and Staveland, L.E. (1988) 'Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research', *Advances in Psychology*, 52(C), pp. 139–183. Available at: https://doi.org/10.1016/S0166-4115(08)62386-9.

Hartstein, A.J. *et al.* (2022) 'Virtual Reality Instructional Design in Orthopedic Physical Therapy Education: A Mixed-Methods Usability Test', *Simulation and Gaming*, 53(2), pp. 111–134. Available at: https://doi.org/10.1177/10468781211073646.

Health Education England (2020) *Enhancing education, clinical practice and staff wellbeing. A national vision for the role of simulation and immersive learning technologies in health and care*. Available at: https://www.hee.nhs.uk/sites/default/files/documents/National Strategic Vision of Sim in Health and Care.pdf.

Health Scholars (2022) *Virtual Reality Clinical Training | Clinicians | Health Scholars*. Available at: https://www.healthscholars.com/.

Hellum, O., Kersten-Oertel, M. and Xiao, Y. (2023) ‘Assessment of user-interaction strategies for neurosurgical data navigation and annotation in virtual reality’, *Virtual Reality*, 27(2), pp. 1345–1355. Available at: https://doi.org/10.1007/s10055-022-00740-5.

Hsieh, M.-J. *et al.* (2015) ‘The effects of different retraining frequency of simulation-based short advanced life support training on health professionals’, *Resuscitation*, 96, p. 90. Available at: https://doi.org/10.1016/j.resuscitation.2015.09.212.

HTC Corporation (2019) *VIVE Pro Eye Launches Today In North America, Setting A New Standard For Enterprise Virtual Reality*. Available at: https://www.vive.com/us/newsroom/2019-06-06/.

Hudson, I. and Hurter, J. (2016) ‘Avatar Types Matter: Review of Avatar Literature for Performance Purposes’, in, pp. 14–21. Available at: https://doi.org/10.1007/978-3-319-39907-2_2.

i3 Simulations (2019) *Resuscitation VR*, *Oculus*. Available at: https://www.oculus.com/experiences/go/2016087265167606/.

i3 Simulations (2022) *i3 Simulations • Innovative integrated technology solutions*. Available at: https://i3simulations.com.

Iliopoulos, E. *et al.* (2018) ‘Correlation between workplace learning and job satisfaction of NHS healthcare professionals’, *British Journal of Healthcare Management*, 24(5), pp. 226–233. Available at: https://doi.org/10.12968/bjhc.2018.24.5.226.

inciteVR (2020) *NP Skills*. Available at: https://incitevr.com/index.php/solutions/healthcare/np-skills-labs.

Ingrassia, P.L. *et al.* (2020) ‘Augmented reality learning environment for basic life support and defibrillation training: Usability study’, *Journal of Medical Internet Research*, 22(5), pp. 1–8. Available at: https://doi.org/10.2196/14910.

International Organization for Standardization (2018) *Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts*. Available at: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en.

Issleib, M. *et al.* (2021) ‘Virtual reality as a teaching method for resuscitation training in undergraduate first year medical students: a randomized controlled trial’, *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, 29(1), p. 27. Available at: https://doi.org/10.1186/s13049-021-00836-y.

Jensen, L. and Konradsen, F. (2017) ‘A review of the use of virtual reality head-mounted displays in education and training’, *Education and Information Technologies*, pp. 1–15. Available at:

https://doi.org/10.1007/s10639-017-9676-0.

Jerald, J. (2016) *The VR Book: Human-Centered Design for Virtual Reality*. Edited by M.T. Özsu and J.C. Hart. ACM Press. Available at: https://doi.org/10.1145/2792790.

Jones, S.E. (2008) *The Meaning of Video Games: Gaming and Textual Strategies*. Routledge.

Kang, H.J., Shin, J. and Ponto, K. (2020) 'A Comparative Analysis of 3D User Interaction: How to Move Virtual Objects in Mixed Reality', in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, pp. 275–284. Available at: https://doi.org/10.1109/VR46266.2020.00047.

Kardong-Edgren, S. *et al.* (2019) 'Evaluating the Usability of a Second-Generation Virtual Reality Game for Refreshing Sterile Urinary Catheterization Skills', *Nurse Educator*, 44(3), pp. 137–141. Available at: https://doi.org/10.1097/NNE.0000000000000570.

Katz, D. *et al.* (2020) 'Utilization of a Voice-Based Virtual Reality Advanced Cardiac Life Support Team Leader Refresher: Prospective Observational Study', *Journal of Medical Internet Research*, 22(3), p. e17425. Available at: https://doi.org/10.2196/17425.

Kockwelp, P. *et al.* (2022) 'Towards VR Simulation-Based Training in Brain Death Determination', in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, pp. 287–292. Available at: https://doi.org/10.1109/VRW55335.2022.00065.

Kyaw, B.M. *et al.* (2019) 'Virtual Reality for Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration', *Journal of Medical Internet Research*, 21(1), p. e12959. Available at: https://doi.org/10.2196/12959.

Latham, K. *et al.* (2019) 'A Review on the Development of a Virtual Reality Learning Environment for Medical Simulation and Training', *Proceedings - 2019 International Conference on Applications and Systems of Visual Paradigms, VISUAL 2019*, (July), pp. 1–5.

Lee, D.K. *et al.* (2022) 'Development of an Extended Reality Simulator for Basic Life Support Training', *IEEE Journal of Translational Engineering in Health and Medicine*, 10, pp. 1–7. Available at: https://doi.org/10.1109/JTEHM.2022.3152365.

Lerner, D. *et al.* (2020) 'An Immersive Multi-User Virtual Reality for Emergency Simulation Training: Usability Study', *JMIR Serious Games*, 8(3), p. e18822. Available at: https://doi.org/10.2196/18822.

Lewis, J.R. and Sauro, J. (2012) *Quantifying the user experience: Practical statistics for user research*. Elsevier.

Liyanage, S.U. *et al.* (2019) 'Towards an Affordable Virtual Reality Solution for Cardiopulmonary Resuscitation Training', in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, pp. 1054–1055. Available at: https://doi.org/10.1109/VR.2019.8798159.

Lombardo, R. *et al.* (2022) 'Ready Medic One: A Feasibility Study of a Semi-Autonomous Virtual Reality Trauma Simulator', *Frontiers in Virtual Reality*, 2. Available at: https://doi.org/10.3389/frvir.2021.719656.

Lucid Reality Labs (2020) *Our XR Technology Projects Portfolio | Lucid Reality Labs*. Available at: https://lucidrealitylabs.com/portfolio.

Lumeto (2021) *Healthcare | Lumeto*. Available at: https://lumeto.com/healthcare.

Lungu, A.J. *et al.* (2021) 'A review on the applications of virtual reality, augmented reality and mixed reality in surgical simulation: an extension to different kinds of surgery', *Expert Review of Medical Devices*, 18(1), pp. 47–62. Available at: https://doi.org/10.1080/17434440.2021.1860750.

Mallik, R. *et al.* (2022) 'Exploring the Role of Virtual Reality to Support Clinical Diabetes Training—A Pilot Study', *Journal of Diabetes Science and Technology*, 16(4), pp. 844–851. Available at: https://doi.org/10.1177/19322968211027847.

McFetrich, J. (2006) 'A structured literature review on the use of high fidelity patient simulators for teaching in emergency medicine', *Emergency Medicine Journal*, 23(7), pp. 509–511. Available at: https://doi.org/10.1136/emj.2005.030544.

McGrath, J.L. *et al.* (2018) 'Using Virtual Reality Simulation Environments to Assess Competence for Emergency Medicine Learners', *Academic Emergency Medicine*. Edited by R. Fernandez, 25(2), pp. 186–195. Available at: https://doi.org/10.1111/acem.13308.

Meta (2020) *Meta Quest 2: Immersive All-in-One VR Headset*. Available at: https://www.meta.com/gb/quest/products/quest-2/.

Meta (2022) *Meta Quest Pro: Built with Privacy in Mind*. Available at: https://www.meta.com/en-gb/blog/quest/meta-quest-pro-privacy/.

Mikhailenko, M., Maksimenko, N. and Kurushkin, M. (2022) 'Eye-Tracking in Immersive Virtual Reality for Education: A Review of the Current Progress and Applications', *Frontiers in Education*, 7(March). Available at: https://doi.org/10.3389/feduc.2022.697032.

Miller, N., Willemsen, P. and Feyen, R. (2018) 'Comparing Interface Affordances for Controlling a Push Broom in VR', *25th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2018 - Proceedings*, pp. 635–636. Available at: https://doi.org/10.1109/VR.2018.8446510.

Moglia, A. *et al.* (2016) 'A Systematic Review of Virtual Reality Simulators for Robot-assisted Surgery', *European Urology*, 69(6), pp. 1065–1080. Available at: https://doi.org/10.1016/j.eururo.2015.09.021.

Moll-Khosrawi, P. *et al.* (2022) 'Virtual reality as a teaching method for resuscitation training in undergraduate first year medical students during COVID-19 pandemic: a randomised controlled trial', *BMC Medical Education*, 22(1), p. 483. Available at: https://doi.org/10.1186/s12909-022-

03533-1.

Muresan, A., McIntosh, J. and Hornbæk, K. (2023) 'Using Feedforward to Reveal Interaction Possibilities in Virtual Reality', *ACM Transactions on Computer-Human Interaction*, pp. 1–47. Available at: https://doi.org/10.1145/3603623.

Ng, H.W. *et al.* (2022) 'Real-Time Spoken Language Understanding for Orthopedic Clinical Training in Virtual Reality', in, pp. 640–646. Available at: https://doi.org/10.1007/978-3-031-11644-5_61.

Norman, D.A. (2013) *The Design of Everyday Things*. MIT Press. Available at: https://muse.jhu.edu/content/crossref/journals/technology_and_culture/v056/56.3.tenner.html.

Ntokos, K. (2019) 'Techniques on Multiplatform Movement and Interaction Systems in a Virtual Reality Context for Games', in *Cases on Immersive Virtual Reality Techniques*, pp. 199–216. Available at: https://doi.org/10.4018/978-1-5225-5912-2.ch009.

Oculus (2020) *Oculus Utilities for Unity*.

Oliveira, T.G.P. de *et al.* (2022) 'Involvement of companions in patient safety in pediatric and neonatal units: scope review', *Revista Brasileira de Enfermagem*, 75(3), p. e20210504. Available at: https://doi.org/10.1590/0034-7167-2021-0504.

Oxford Medical Simulation (2022) *Oxford Medical Simulation - Virtual Reality Healthcare Training*. Available at: https://oxfordmedicalsimulation.com/.

Pan, Z. *et al.* (2010) 'A real-time multi-cue hand tracking algorithm based on computer vision', *Proceedings - IEEE Virtual Reality*, (April 2010), pp. 219–222. Available at: https://doi.org/10.1109/VR.2010.5444787.

Pantelidis, P. *et al.* (2018) 'Virtual and Augmented Reality in Medical Education', in *Medical and Surgical Education - Past, Present and Future*. InTech, p. 13. Available at: https://doi.org/10.5772/intechopen.71963.

Parham, G. *et al.* (2019) 'Creating a low-cost virtual reality surgical simulation to increase surgical oncology capacity and capability', *ecancermedicalscience*, 13, pp. 1–11. Available at: https://doi.org/10.3332/ecancer.2019.910.

Peng, C. *et al.* (2021) 'Rapid Cycle Deliberate Practice in Virtual Reality: Teaching Transvenous Pacemaker Insertion to Emergency Medicine Residents', *Cureus*, 13(10), p. e18503. Available at: https://doi.org/10.7759/cureus.18503.

Perron, J.E. *et al.* (2021) 'Resuscitating Cardiopulmonary Resuscitation Training in a Virtual Reality: Prospective Interventional Study', *Journal of Medical Internet Research*, 23(7), p. e22920. Available at: https://doi.org/10.2196/22920.

Peterson, E., Porter, M. and Calhoun, A. (2020) 'Mixed-Reality Simulation for a Pediatric Transport Team: A Pilot Study', *Air Medical Journal*, 39(3), pp. 173–177. Available at: https://doi.org/10.1016/j.amj.2020.03.001.

Pfeuffer, K. *et al.* (2020) 'Empirical Evaluation of Gaze-enhanced Menus in Virtual Reality', in *26th ACM Symposium on Virtual Reality Software and Technology*. New York, NY, USA: ACM, pp. 1–11. Available at: https://doi.org/10.1145/3385956.3418962.

Putnam, E.M. *et al.* (2021) 'Virtual reality simulation for critical pediatric airway management training.', *Journal of Clinical and Translational Research*, 7(1), pp. 93–99. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC8177020.

Radianti, J. *et al.* (2020) 'A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda', *Computers & Education*, 147(December 2019), p. 103778. Available at: https://doi.org/10.1016/j.compedu.2019.103778.

Ralston, B.H. *et al.* (2021) 'Use of Virtual Reality for Pediatric Cardiac Critical Care Simulation', *Cureus*, 13(6). Available at: https://doi.org/10.7759/cureus.15856.

Rangarajan, K., Davis, H. and Pucher, P.H. (2020) 'Systematic Review of Virtual Haptics in Surgical Simulation: A Valid Educational Tool?', *Journal of Surgical Education*, 77(2), pp. 337–347. Available at: https://doi.org/10.1016/j.jsurg.2019.09.006.

Rauschnabel, P.A. *et al.* (2022) 'What is XR? Towards a Framework for Augmented and Virtual Reality', *Computers in Human Behavior*, 133(March), p. 107289. Available at: https://doi.org/10.1016/j.chb.2022.107289.

Reason, J. (1990) *Human Error*. Cambridge University Press.

Sapkaroski, D., Mundy, M. and Dimmock, M.R. (2022) 'Immersive virtual reality simulated learning environment versus role-play for empathic clinical communication training', *Journal of Medical Radiation Sciences*, 69(1), pp. 56–65. Available at: https://doi.org/10.1002/jmrs.555.

Scafà, M. *et al.* (2020) 'Assessment of Students' Cognitive Conditions in Medical Simulation Training: A Review Study', in *AHFE 2019: Advances in Human Factors and Simulation*, pp. 224–233. Available at: https://doi.org/10.1007/978-3-030-20148-7_21.

Schild, J., Misztal, S., *et al.* (2018) 'Applying Multi-User Virtual Reality to Collaborative Medical Training', in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, pp. 775–776. Available at: https://doi.org/10.1109/VR.2018.8446160.

Schild, J., Lerner, D., *et al.* (2018) 'EPICSAVE — Enhancing vocational training for paramedics with multi-user virtual reality', in *2018 IEEE 6th International Conference on Serious Games and Applications for Health (SeGAH)*. IEEE, pp. 1–8. Available at: https://doi.org/10.1109/SeGAH.2018.8401353.

Schild, J., Elsenbast, C. and Carbonell, G. (2021) 'ViTAWiN - Developing Multiprofessional Medical Emergency Training with Mixed Reality', in *2021 IEEE 9th International Conference on Serious Games and Applications for Health(SeGAH)*. IEEE, pp. 1–9. Available at: https://doi.org/10.1109/SEGAH52098.2021.9551890.

Schlattman, M. and Klein, R. (2007) 'Simultaneous 4 gestures 6 DOF real-time two-hand tracking without any markers', *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST*, 1(212), pp. 39–42. Available at: https://doi.org/10.1145/1315184.1315188.

Semeraro, F. *et al.* (2009) 'Virtual reality enhanced mannequin (VREM) that is well received by resuscitation experts', *Resuscitation*, 80(4), pp. 489–492. Available at: https://doi.org/10.1016/j.resuscitation.2008.12.016.

Semeraro, F., Ristagno, G., Giulini, G., Kayal, J.S., *et al.* (2019) 'Back to reality: A new blended pilot course of Basic Life Support with Virtual Reality', *Resuscitation*, 138(0), pp. 18–19. Available at: https://doi.org/10.1016/j.resuscitation.2019.02.034.

Semeraro, F., Ristagno, G., Giulini, G., Gnudi, T., *et al.* (2019) 'Virtual reality cardiopulmonary resuscitation (CPR): Comparison with a standard CPR training mannequin', *Resuscitation*, 135, pp. 234–235. Available at: https://doi.org/10.1016/j.resuscitation.2018.12.016.

Shum, L.C. *et al.* (2023) *Personalised Learning through Context-Based Adaptation in the Serious Games with Gating Mechanism*, *Education and Information Technologies*. Springer US. Available at: https://doi.org/10.1007/s10639-023-11695-8.

SimX (2021) *Virtual Reality Medical Simulation | SimX – The world's best Virtual reality medical training software and nursing simulation training platform*. Available at: https://www.simxvr.com.

Sivakumar, P. *et al.* (2021) 'Grasping performance depends upon the richness of hand feedback', *Experimental Brain Research*, 239(3), pp. 835–846. Available at: https://doi.org/10.1007/s00221-020-06025-0.

Society for Simulation in Healthcare (2022) *Corporate Roundtable*. Available at: https://www.ssih.org/Membership-Engagement/Corporate-Roundtable (Accessed: 3 March 2022).

Strazdins, G. *et al.* (2017) 'Virtual reality using gesture recognition for deck operation training', in *OCEANS 2017 - Aberdeen*. IEEE, pp. 1–6. Available at: https://doi.org/10.1109/OCEANSE.2017.8084584.

Team PCS (2022) *PCS Patient Communication Simulators*. Available at: https://www.pcs.ai/spark/.

Tu, H. *et al.* (2019) 'Crossing-Based Selection with Virtual Reality Head-Mounted Displays', in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, pp. 1–14. Available at: https://doi.org/10.1145/3290605.3300848.

UbiSim (2022) *The Virtual Reality Training Platform for Nursing*. Available at: https://ubisimvr.com.

Unity Technologies (2018) 'Unity'.

Valve Corporation (2019) *Valve Index® - Upgrade your experience*. Available at: https://www.valvesoftware.com/en/index.

Vaughan, N. *et al.* (2016) 'A review of virtual reality based training simulators for orthopaedic surgery', *Medical Engineering & Physics*, 38(2), pp. 59–71. Available at: https://doi.org/10.1016/j.medengphy.2015.11.021.

Vaughan, N., Gabrys, B. and Dubey, V.N. (2016) 'An overview of self-adaptive technologies within virtual reality training', *Computer Science Review*, 22, pp. 65–87. Available at: https://doi.org/10.1016/j.cosrev.2016.09.001.

Virtual Reality and Augmented Reality Wiki (2017) *Standing VR*, *Virtual Reality and Augmented Reality Wiki*.

Virtual Reality Society (2020) *History of Virtual Reality*.

Voigt-Antons, J.N. *et al.* (2020) 'Influence of Hand Tracking as a Way of Interaction in Virtual Reality on User Experience', *2020 12th International Conference on Quality of Multimedia Experience, QoMEX 2020*, pp. 2020–2023. Available at: https://doi.org/10.1109/QoMEX48832.2020.9123085.

VRpatients (2022) *Virtual Simulation Education for Emergency Responders & Nurses | VRpatients*. Available at: https://vrpatients.com.

Wang, R.Y. and Popović, J. (2009) 'Real-time hand-tracking with a color glove', *ACM Transactions on Graphics*, 28(3), pp. 1–8. Available at: https://doi.org/10.1145/1531326.1531369.

WaterWell LLC (2022) *Healthcare Simulation Vendors*. Available at: https://www.healthysimulation.com/medical-simulation/vendors/ (Accessed: 3 March 2022).

Weber, S., Weibel, D. and Mast, F.W. (2021) 'How to Get There When You Are There Already? Defining Presence in Virtual Reality and the Importance of Perceived Realism', *Frontiers in Psychology*, 12(May), pp. 1–10. Available at: https://doi.org/10.3389/fpsyg.2021.628298.

Wei, X., Jin, X. and Fan, M. (2022) 'Communication in Immersive Social Virtual Reality: A Systematic Review of 10 Years' Studies', *Educational Technology and Society*, 24(2), pp. 93–106. Available at: http://arxiv.org/abs/2210.01365.

Widjojo, E.A., Chinthammit, W. and Engelke, U. (2017) 'Virtual Reality-based Human-Data Interaction', in *2017 International Symposium on Big Data Visual Analytics (BDVA)*. Available at:

https://ieeexplore.ieee.org/document/8114627?reload=true.

Wong, M.A.M.E. *et al.* (2018) ‘Clinical instructors’ perceptions of virtual reality in health professionals’ cardiopulmonary resuscitation education’, *SAGE Open Medicine*, 6, p. 205031211879960. Available at: https://doi.org/10.1177/2050312118799602.

World Health Organization (2022) *Working for Health 2022-2030 Action Plan*.

Wozniak, P. *et al.* (2016) ‘Possible applications of the LEAP motion controller for more interactive simulated experiments in augmented or virtual reality’, in G.G. Gregory (ed.), p. 99460P. Available at: https://doi.org/10.1117/12.2237673.

Wu, M.-L., Chao, L.-F. and Xiao, X. (2022) ‘A pediatric seizure management virtual reality simulator for nursing students: A quasi-experimental design’, *Nurse Education Today*, 119, p. 105550. Available at: https://doi.org/10.1016/j.nedt.2022.105550.

Yang, C.-H. *et al.* (2020) ‘Immersive Virtual Reality-Based Cardiopulmonary Resuscitation Interactive Learning Support System’, *IEEE Access*, 8, pp. 120870–120880. Available at: https://doi.org/10.1109/ACCESS.2020.3006280.

Yang, S.-Y. and Oh, Y.-H. (2022) ‘The effects of neonatal resuscitation gamification program using immersive virtual reality: A quasi-experimental study’, *Nurse Education Today*, 117, p. 105464. Available at: https://doi.org/10.1016/j.nedt.2022.105464.

Yildirim, C. (2022) ‘Point and Select: Effects of Multimodal Feedback on Text Entry Performance in Virtual Reality’, *International Journal of Human–Computer Interaction*, 0(0), pp. 1–15. Available at: https://doi.org/10.1080/10447318.2022.2107330.

Zackoff, M.W. *et al.* (2021) ‘Establishing Objective Measures of Clinical Competence in Undergraduate Medical Education Through Immersive Virtual Reality’, *Academic Pediatrics*, 21(3), pp. 575–579. Available at: https://doi.org/10.1016/j.acap.2020.10.010.

Zhang, J. *et al.* (2023) ‘Understanding the effects of hand design on embodiment in virtual reality’, *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 37, p. e10. Available at: https://doi.org/10.1017/S0890060423000045.

Zhang, X. *et al.* (2017) ‘How virtual reality affects perceived learning effectiveness: a task–technology fit perspective’, *Behaviour & Information Technology*, 3001(June), pp. 1–9. Available at: https://doi.org/10.1080/0144929X.2016.1268647.

Appendix A – Clinical VR Training Literature Review Table

H/W = Hardware Used   CT = Control Trial   U# = Number of Users

Title	H/W	Topic	CT	#	Quant.	Qual.	Frictions
A mixed-methods feasibility study to assess the acceptability and applicability of immersive virtual reality sepsis game as an adjunct to nursing education (Adhikari *et al.*, 2021)	Oculus Rift	Nursing	N	19		26.1% confidence increase 23.4% anxiety decrease	Lack of verbal actions VR hardware
A pediatric seizure management virtual reality simulator for nursing students: A quasi-experimental design (Wu, Chao and Xiao, 2022)	Oculus Rift	Resus*	N	105	Higher post-test scores (t = 5.05)	77% VR acceptance 80% agree usefulness score	VR controls
An Immersive Multi-User Virtual Reality for Emergency Simulation Training: Usability Study (Lerner *et al.*, 2020)	HTC Vive	Resus*	N	18		87% perceived effectiveness 66% usability	VR controls
Application of VR Technology to the Training of Paramedics (Boros *et al.*, 2022)	Meta Quest	Para†	N	120	97.5% pass rate	99% VR interest	VR hardware
Back to reality: A new blended pilot course of Basic Life Support with Virtual Reality (Semeraro, Ristagno, Giulini, Kayal, *et al.*, 2019)	HTC Vive	BLS	Y	22		100% VR interest 72% perceived ease of use	
Clinical instructors' perceptions of virtual reality in health professionals' cardiopulmonary resuscitation education (Wong *et al.*, 2018)	HTC Vive	CPR	N	30		High fidelity	

Comparisons of Stress Physiology of Providers in Real-Life Resuscitations and Virtual Reality–Simulated Resuscitations (Chang *et al.*, 2019)	Oculus Rift	Resus*	Y	16	Lower heart rate change in VR (47% lower)	26% increase in perceived task load compared to real events	VR controls
Development and Considerations for Virtual Reality Simulations for Resuscitation Training and Stress Inoculation (Chang *et al.*, 2020)	Oculus Rift	Resus*	N	34	Increase in cortisol levels for residents (+0.07 µg/dL)		VR controls
Development and evaluation of a trauma decision-making simulator in Oculus virtual reality (Harrington *et al.*, 2018)	Gear VR	Trauma	N	30	78.9% pass rate	81% perceived usefulness	VR controls
Development of an Extended Reality Simulator for Basic Life Support Training (Lee *et al.*, 2022)	HTC Vive	BLS	N	16		93% perceived ease of use 92% effectiveness	
EPICSAVE — Enhancing vocational training for paramedics with multi-user virtual reality (Schild, Lerner, *et al.*, 2018)	Oculus Rift	Para†	N	24		64% usability 76% perception score	VR hardware
Establishing Objective Measures of Clinical Competence in Undergraduate Medical Education Through Immersive Virtual Reality (Zackoff *et al.*, 2021)	Oculus Rift	Resusc*	N	26	78% average score reliability		
Evaluating the Usability of a Second-Generation Virtual Reality Game for Refreshing Sterile Urinary Catheterization Skills (Kardong-Edgren *et al.*, 2019)	Oculus Rift	Nursing	N	31		64% usability 75% reflection score	VR hardware VR controls

Exploring the Role of Virtual Reality to Support Clinical Diabetes Training—A Pilot Study (Mallik *et al.*, 2022)	Oculus Rift	Diagnosis	N	39		72% confidence increase	
Immersive virtual reality simulated learning environment versus role-play for empathic clinical communication training (Sapkaroski, Mundy and Dimmock, 2022)	Oculus Rift	Comm**	Y	79		5% self-efficiency increase	
Immersive Virtual Reality-Based Cardiopulmonary Resuscitation Interactive Learning Support System (Yang *et al.*, 2020)	Google VR	CPR	N	100	85% pass rate for trainees	87% satisfaction 85% perceived usability	
Is individual practice in an immersive and interactive virtual reality application non-inferior to practicing with traditional equipment in learning systematic clinical observation? A randomized controlled trial (Berg and Steinsbekk, 2020)	Oculus Rift Meta Quest	Observation	Y	289	2.3% VR test score increase	46% higher VR acceptance 79.7% VR usability	
Rapid Cycle Deliberate Practice in Virtual Reality: Teaching Transvenous Pacemaker Insertion to Emergency Medicine Residents (Peng *et al.*, 2021)	Meta Quest	ACLS	N	16	19% test scores increase	40% perceived preparedness increase	
Ready Medic One: A Feasibility Study of a Semi-Autonomous Virtual Reality Trauma Simulator (Lombardo *et al.*, 2022)	Windows Mixed Reality	Trauma	N	17	30% controls difficulty	84% perceived usefulness	VR controls VR hardware

					29% hardware difficulty		Lack of verbal actions
Resuscitating Cardiopulmonary Resuscitation Training in a Virtual Reality: Prospective Interventional Study (Perron *et al.*, 2021)	Oculus Rift	Resus*	N	26		70% perceived usefulness 30% perceived usability	VR controls
The effects of neonatal resuscitation gamification program using immersive virtual reality: A quasi-experimental study (Yang and Oh, 2022)	Oculus Rift	Resus*	Y	83	34% test score increase 13% problem-solving increase	24.7% confidence increase	VR controls
Towards an Affordable Virtual Reality Solution for Cardiopulmonary Resuscitation Training (Liyanage *et al.*, 2019)	HTC Vive	CPR	Y	69		Expert issues with interaction High acceptance	VR hardware VR controls
Towards VR Simulation-Based Training in Brain Death Determination (Kockwelp *et al.*, 2022)	Valve Index	Diagnosis	N	3		High acceptance Usability issues	VR controls
Trauma bay virtual reality—A game changer for ATLS instruction and assessment (Colonna *et al.*, 2022)	Meta Quest	Trauma	N	31	High reliability	82% acceptance 79% simulation satisfaction	Lack of verbal actions VR controls
Use of Virtual Reality for Pediatric Cardiac Critical Care Simulation (Ralston *et al.*, 2021)	Meta Quest	ACLS	N	6		83% perceived usefulness	VR hardware

						63% perceived usability	
Utilization of a Voice-Based Virtual Reality Advanced Cardiac Life Support Team Leader Refresher: Prospective Observational Study (Katz *et al.*, 2020)	Windows Mixed Reality	ACLS	Y	25	52% shorter session length 54% lower cost 35% lower test score	80% lower task load 8% lower perceived usefulness	VR controls
Virtual reality as a teaching method for resuscitation training in undergraduate first year medical students during COVID-19 pandemic: a randomised controlled trial (Moll-Khosrawi *et al.*, 2022)	HTC Vive	Resus*	Y	88	53% error reduction	45% confidence increase	
Virtual reality as a teaching method for resuscitation training in undergraduate first year medical students: a randomized controlled trial (Issleib *et al.*, 2021)	HTC Vive	Resus*	Y	160	55% learning increase	93% usability	
Virtual reality cardiopulmonary resuscitation (CPR): Comparison with a standard CPR training mannequin (Semeraro, Ristagno, Giulini, Gnudi, *et al.*, 2019)	HTC Vive	CPR	Y	43	High score parity		
Virtual reality simulation for critical pediatric airway management training (Putnam *et al.*, 2021)	HoloLens	Airway	N	41	19% test score increase	77% usability	VR controls VR hardware

Virtual reality simulation for learning wound dressing: Acceptance and usability (Choi, 2022)	HTC Vive	Nursing	Y	30	High score parity	5.91% higher acceptance	VR controls
Virtual Reality Simulation Technology for Cardiopulmonary Resuscitation Training: An Innovative Hybrid System With Haptic Feedback (Almousa *et al.*, 2019)	HTC Vive	CPR	N	20		85% confidence	VR controls
Virtual Reality vs. High-Fidelity Mannequin-Based Simulation: A Pilot Randomized Trial Evaluating Learner Performance (Abulfaraj *et al.*, 2021)	Oculus Rift	Resus*	Y	42	High score parity	10% confidence increase 77% usability	
Virtual reality-based neurological examination teaching tool (VRNET) versus standardized patient in teaching neurological examinations for the medical students: a randomized, single-blind study (Han *et al.*, 2021)	Oculus Rift	Diagnosis	Y	98	11% higher test score	Satisfaction parity	Lack of verbal actions
ViTAWiN - Developing Multiprofessional Medical Emergency Training with Mixed Reality (Schild, Elsenbast and Carbonell, 2021)	HTC Vive	Para†	N	28		72.1% usability	VR hardware VR controls Lack of verbal actions

** Resuscitation ** Communication † Paramedic*

Appendix B – NASA-TLX and SUS User Trial Scores

Score	Scenario #1	Scenario #2	Change	Average
TLX Mental[1]	57.22 ± 22.37	58.13 ± 24.99	+ 2%	57.67 ± 23.65
TLX Physical[1]	13.89 ± 6.98	13.75 ± 6.50	- 1%	13.82 ± 6.76
TLX Temporal[1]	57.22 ± 26.89	61.88 ± 28.28	+ 8%	59.55 ± 27.65
TLX Performance[2]	43.33 ± 26.03	61.88 ± 26.33	+ 43%	52.60 ± 27.76
TLX Effort[1]	63.89 ± 13.29	59.38 ± 14.24	- 7%	61.63 ± 13.93
TLX Frustration[1]	57.78 ± 25.29	58.75 ± 27.92	+ 2%	58.26 ± 26.57
Workload (Raw TLX)[1]	51.11 ± 8.74	48.33 ± 12.92	- 5%	49.72 ± 10.99
SUS Frequency[2,3]	4.38 ± 0.48	4.43 ± 0.73	+ 1%	4.40 ± 0.61
SUS Complexity[1,3]	2.25 ± 0.97	2.00 ± 0.76	- 11%	2.13 ± 0.88
SUS Easy[2,3]	3.88 ± 0.78	3.86 ± 0.99	- 0.5%	3.87 ± 0.88
SUS Support[1,3]	3.13 ± 0.93	2.86 ± 0.99	- 9%	2.99 ± 0.97
SUS Integrated[2,3]	3.88 ± 0.33	3.71 ± 0.45	- 4%	3.79 ± 0.40
SUS Inconsistency[1,3]	2.75 ± 0.66	2.43 ± 0.49	- 12%	2.59 ± 0.61
SUS Learn Quickly[2,3]	3.88 ± 0.60	3.86 ± 0.83	- 0.5%	3.87 ± 0.72
SUS Awkward[1,3]	2.13 ± 1.05	2.00 ± 1.07	- 6%	2.06 ± 1.06
SUS Confident[2,3]	3.50 ± 1.12	3.57 ± 1.05	+ 2%	3.54 ± 1.09
SUS Learn Before[1,3]	2.25 ± 1.09	1.86 ± 1.12	- 17%	2.05 ± 1.12
System Usability Score[2]	67.50 ± 14.68	70.71 ± 15.22	+ 5%	69.11 ± 15.05
Bangor Rating[2]	5.63 ± 0.70	5.57 ± 0.73	- 2%	5.60 ± 0.71

[1]Lower is better; [2]Higher is better; [3]1=strongly disagree, 5=strongly agree; Values are presented as mean ± SD.

Appendix C – User Error Results

	CI[1]	SM[2]	EA[3]	UMR[4]	UAR[5]	UPR[6]	UAO[7]
Count S1	6 (67%[8])	3 (33%)	2 (22%)	1 (11%)	4 (44%)	1 (11%)	3 (33%)
Count S2	5 (56%)	1 (11%)	2 (22%)	5 (56%)	3 (33%)	1 (11%)	1 (11%)
Count Total	11 (61%)	4 (22%)	4 (22%)	6 (33%)	7 (39%)	2 (11%)	4 (22%)
Diff Count	-17%	-67%	0%	400%	-25%	0%	-67%
Sum S1	8	5	2	3	4	2	3
Sum S2	5	2	2	7	3	3	1
Sum Total	13	7	4	10	7	5	4
Diff Sum	-38%	-67%	0%	133%	-25%	50%	-67%
Avg Per Sc.	1.18	1.75	1.00	1.67	1.00	2.50	1.00

[1]Controller Issue; [2]State Misidentify; [3]Environment Awareness; [4]Uncoded Medication Request; [5]Uncoded Airway Request; [6]Uncoded Procedure Request; [7]Uncoded Action Order; [8]Percentage of scenario sessions in which error occurred.

Appendix D – Market Analysis References

Organisation	Relevant Product(s)	Data	Reference URLs
E Learning Design Center (2022)	MedVR Education	Website Simulation Videos	medvr.education vimeo.com/user106612794
GigXR (2022)	HoloScenarios	Website Wiki Simulation Videos	gigxr.com help.gigxr.com/knowledge vimeo.com/user106277094
Health Scholars (2022)	Hospital VR, EMS VR	Website Wiki Simulation Videos	healthscholars.com healthscholars.com/knowledge-base youtube.com/@healthscholars406
i3 Simulations (2022)	Resuscitation VR	Full Application	i3simulations.com/emergency-medicine
inciteVR (2020)	Medical Assisting Clinical Suite	Website Simulation Videos	incitevr.com youtube.com/@inciteVR
Lucid Reality Labs (2020)	Medical VR Intubation VR Hysteroscopy Multiplayer	Website Simulation Videos	lucidrealitylabs.com youtube.com/@lucidrealitylabs6378
Lumeto (2021)	Involve XR	Website Simulation Videos	lumeto.com/healthcare youtube.com/@lumeto5283
Oxford Medical Simulation (2022)	OMS Nursing OMS Medical	Website Wiki Simulation Videos	oxfordmedicalsimulation.com intercom.help/oxford-medical-simulation youtu.be/CV4JW7GN9Yg

PCS (2022)	PCS Spark	Website Wiki Simulation Videos	pcs.ai/spark help.pcs.ai vimeo.com/pcsna
SimX (2021)	SimX EMS	Website Simulation Videos Application Demo	simxvr.com youtube.com/@SimXVR oculus.com/experiences/quest/3422065107904151
UbiSim (2022)	UbiSim VR	Website Simulation Videos	ubisimvr.com youtube.com/@ubisim1669
VRpatients (2022)	VRpatients for Nursing	Website Simulation Videos Application Demo	vrpatients.com youtube.com/@vrpatients oculus.com/experiences/quest/3756091017802002

Appendix E – Clinical Skills Representation

ID	Clinical Action	No. Support	No. Evaluable
1a	Asking Patient Questions	9 (75%)	8 (67%)
1b	Asking Patient Companion Questions	4 (33%)	3 (25%)
1c	Asking Healthcare Personnel Questions	8 (67%)	4 (33%)
2a	Examining Patient	11 (92%)	8 (67%)
2b	Manipulating Patient Physically	2 (17%)	1 (8%)
2c	Note Taking	1 (8%)	0 (0%)
3a	Ordering Medical Tests	9 (75%)	7 (58%)
3b	Ordering Medical Scans	9 (75%)	7 (58%)
3c	Performing Medical Tests	11 (92%)	10 (83%)
4a	Reading Patient Vitals	11 (92%)	2 (17%)
4b	Reading Patient Information	12 (100%)	2 (17%)
4c	Developing an Action Plan	5 (42%)	3 (25%)
5a	Administering Medication	10 (83%)	9 (75%)
5b	Administering Airway Management	10 (83%)	9 (75%)
5c	Administering Medical Equipment	10 (83%)	9 (75%)
6a	Communicating with Patient	7 (58%)	6 (50%)
6b	Communicating with Patient Companion	4 (33%)	3 (25%)
6c	Communicating with Hospital Departments	5 (42%)	5 (42%)
7a	Closed-Loop Communication	6 (50%)	1 (8%)
7b	Giving Instructions	7 (58%)	3 (25%)
7c	Role Assignment	2 (17%)	1 (8%)
8a	Following Protocol	12 (100%)	10 (83%)
8b	Empathetic Communication	5 (42%)	3 (25%)
8c	Risk Analysis	1 (8%)	0 (0%)

Appendix F – Market Analysis Interaction Matrix

M = Multi-user?   S = Supported   E = Evaluable   Action IDs in Table 8 – Clinical Actions.

Name	M	1a	1b	1c	2a	2b	2c	3a	3b	3c	4a	4b	4c	5a	5b	5c	6a	6b	6c	7a	7b	7c	8a	8b	8c
ELDC*	Y	E	E	S	E			S	S	E	E	E	E	E	E	E	E	E	E	S	S		E	E	
GigXR	Y				S			S	S	S	S	S	S	S	S	S				S	S		S	S	
Health Scholars	N	E	E	E				E	E		S	S	E				E	E	E	E	E		E	E	
i3 Simulations	Y	E		E	E	E		E	E	E	S	S	E	E	E	E			E		E	E	E		
inciteVR	N	E		E	E		S		E	E	E	S		E	E	E	E		E				E		
Lucid Reality Labs	N				E					E	S	S		E	E	E							E		
Lumeto	Y			S	S					E	S	S	S	E	E	E				S		S	S		S
OMS**	Y	E	E	E	E			E	E	E	S	E		E	E	E	E	E	E		E		E		
PCS	Y	E			E			E	E	E		S					E						E	E	
SimX	Y	E		S	E			E	E	E	S	S		E	E	E	S			S	S		E		
UbiSim	Y	S	S	S	S	S		E		E	S	S		E	E	E		S		S	S		E	S	
VRpatients	N	E			E			E	E	E	S	S		E	E	E	E						E		

**E Learning Design Center **Oxford Medical Simulation*

Appendix G – Market Products Skills Representation

Organisation	No. Actions Supported	No. Action Evaluable
E Learning Design Center	20 (83%)	15 (63%)
GigXR	14 (58%)	0 (0%)
Health Scholars	15 (63%)	13 (54%)
i3 Simulations	17 (71%)	15 (63%)
inciteVR	14 (58%)	12 (50%)
Lucid Reality Labs	8 (33%)	6 (25%)
Lumeto	13 (54%)	4 (17%)
Oxford Medical Simulation	17 (71%)	16 (67%)
PCS	9 (38%)	8 (33%)
SimX	15 (63%)	9 (38%)
UbiSim	17 (71%)	6 (25%)
VRpatients	12 (50%)	10 (42%)

Appendix H – Design Recommendations

Name	Description
Affordance previews	Use feedforward previews to demonstrate clinical object usages. Explicit previews are suitable for tutorialisation, and 'hidden' previews may be appropriate for in-simulation use.
State information interface	Implement clinical actions for tracking actions such as medication given, and information requests from clinical staff using voice control or conversational interfaces. Voice cues instead of text incur greater information exchange.
Voice control	Implement voice control using Natural Language Processing for intuitive interactions and allowing communication training, as available from contemporary software.
Hand tracking	Implement hand tracking as an alternative control scheme to using VR controllers for intuitive interactions, as available on some contemporary VR hardware. Use object highlights and open-hand gestures, as well as multimodal feedback. Design virtual hands with high trackability but low-medium visual fidelity unless met by interaction fidelity.
Compound selection	Implement compound selection interaction designs for faster interaction cycles, as seen in other non-medical training simulations. Explore novel interaction designs such as worlds-in-miniature depending on virtual object affordance.
Concurrent interaction modalities	Investigate and implement concurrent interaction modalities for closer parity with real-world actions (allowing voice control alongside hand controllers, for example), as seen in some VR clinical simulations.
Manipulating patient physically	Investigate and implement interaction design for manipulating patient physically, using soft body designs already incorporated into surgical training.

Patient companion	Incorporate a patient companion into scenario designs, as an additional factor for information gathering and communication, as seen in real-world practice and some VR clinical simulations.
Handwriting	Investigate and implement interaction design for supporting handwriting, particularly for documentation in simulation.
Note taking	Integrate interfaces for note taking during simulation, using crossing selection and multimodal feedback for VR text input.
Non-Verbal Communication	Design virtual agents to incorporate non-verbal symbolic and natural gestures as part of communication and information exchange.
Eye tracking	Investigate and implement eye tracking to support empathy and observation training, as available on some contemporary VR hardware. Circle Marker is effective visualisation of gaze.