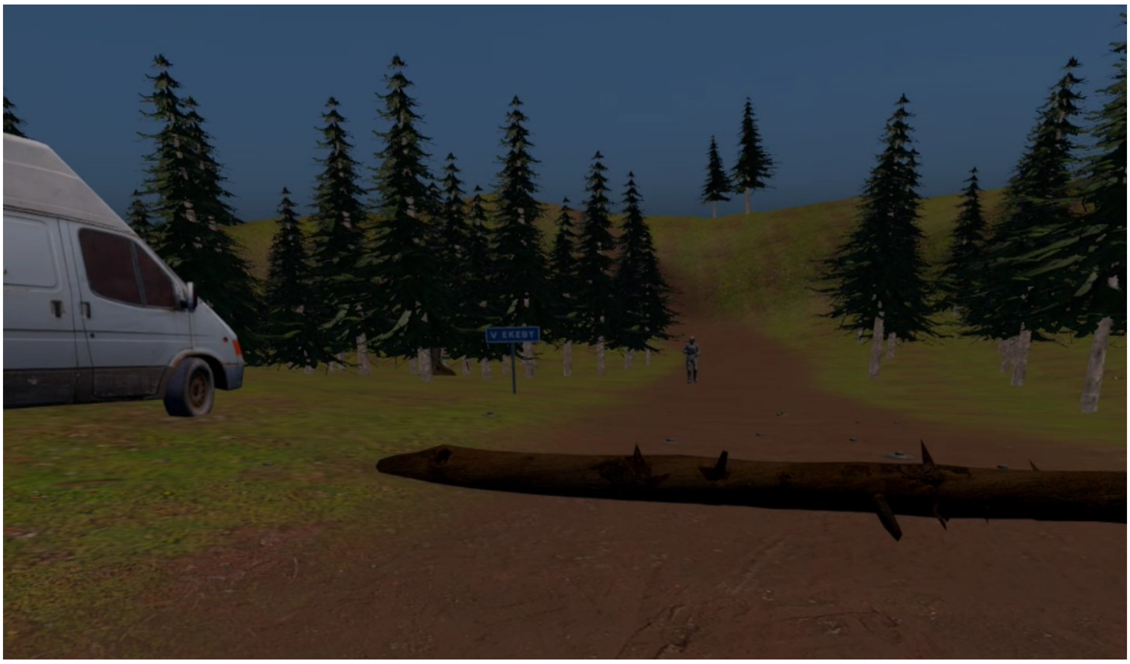
CHALMERS
UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG



Casualty Care Simulator for Standalone VR

Proposing Areas of Consideration when Designing Casualty Care Simulators for Standalone VR Systems

Master's Thesis in Interaction Design and Technologies

HANNES NORDBLOM
JONATHAN STIGSON

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2024
www.chalmers.se

Master's Thesis 2024

Casualty Care Simulator for Standalone VR

Proposing Areas of Consideration when Designing Casualty Care Simulators for Standalone VR Systems

HANNES NORDBLOM
JONATHAN STIGSON

UNIVERSITY OF GOTHENBURG

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
Chalmers University of Technology
University of Gothenburg
Gothenburg, Sweden 2024

Casualty Care Simulator for Standalone VR
Proposing Areas of Consideration when Designing Casualty Care Simulators for Standalone VR Systems
HANNES NORDBLOM
JONATHAN STIGSON

Supervisor: Morten Fjeld, Department of Computer Science and Engineering
Examiner: Staffan Björk, Department of Computer Science and Engineering

Master's Thesis 2024
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Sweden
Telephone +46 31 772 1000

Cover: The final prototype for a standalone VR system set in a small forest town with a dirt road, a van on the left, a fallen tree, and a soldier walking up the road.

Typeset in LaTeX
Gothenburg, Sweden 2024

Casualty Care Simulator for Standalone VR
Proposing Areas of Consideration when Designing Casualty Care Simulators for Standalone VR Systems
HANNES NORDBLOM
JONATHAN STIGSON
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

VR is becoming a prominent tool in many fields. Military applications are no exception and VR is already used extensively in various military fields such as combat training and flight simulation, allowing learners to experience an environment resembling what they would encounter in real life. In this thesis, casualty care is explored by developing a simulator for a standalone VR system that teaches the users what steps to take when a squadmate amputates a limb resulting from a landmine explosion. The prototype was tested in a user study using participants (n = 13) with and without military backgrounds. The results from the quantitative data analysis showed that there was no statistically significant effect on any of the 28 dependent variables, most of which were primarily indicative of performance. The qualitative data analysis, supported by the quantitative data analysis, resulted in four design considerations that one should take into account when designing casualty care simulators for standalone VR systems.

Keywords: casualty care, virtual reality, hmd, standalone vr, military medicine

Acknowledgements

We would like to thank our supervisor Morten Fjeld for supporting us throughout our thesis work and giving valuable input whenever necessary. We would also like to thank Tactisim for giving us the freedom to develop what we wanted and for providing us with the hardware necessary for the thesis to become reality.

Hannes Nordblom & Jonathan Stigson, Gothenburg, May 2024

List of Acronyms

Below is the list of acronyms that have been used throughout the thesis listed in alphabetical order:

API	Application Programming Interface
AR	Augmented Reality
AT	Activity Theory
CLT	Cognitive Load Theory
CPR	Cardiopulmonary Resuscitation
DT	Design Thinking
DV	Dependent Variable
FBS	Failures Before Success
GUI	Graphical User Interface
HCI	Human-Computer Interaction
HMD	Head Mounted Display
IV	Independent Variable
LMIC	Low- and Middle-Income Country
NASA-TLX	NASA Task Load Index
RL	Real Life
RtD	Research through Design
RTLX	Raw Task Load Index
SBT	Scenario-Based Training
SUS	System Usability Scale
SUS-Q	Slater-Usoh-Steed Questionnaire
TAM	Technology Acceptance Model
TCCC	Tactical Combat Casualty Care
TCT	Task Completion Time
UES	User Engagement Scale
UES-SF	User Engagement Scale Short Format
VR	Virtual Reality
WS	Witmer and Singer's Presence Questionnaire
XR	Extended Reality

List of Definitions

Below is the list of definitions that have been used throughout the thesis listed in alphabetical order:

Casualty care	An umbrella term for emergency medical and trauma care in civilian and military contexts.
Casualty care simulator	A system or device used to replicate medical scenarios and provide practical training for healthcare professionals, particularly in emergency medicine and trauma care.
Standalone VR	Virtual reality technology designed to be self-contained, that is running and rendering software without requiring a connection to external hardware.
High-end VR	Advanced virtual reality setups engineered for professionals and typically not available for purchase by hobbyists. These setups also require a tethered computer, extra sensors, and cameras that all cooperate to achieve the highest graphical fidelity possible.
Mobile	A device such as a smartphone or a tablet. Laptops, which typically require a physical keyboard and are less compact, are not included in this definition.
Pre-hospital care	Medical care or procedures administered before a patient reaches a hospital.
User test	A user's participation in the study. Includes interaction with the prototype, filling out questionnaires and taking part in interviews answering questions about their experience with the prototype.
Usability test	The method describing the form of the user tests. In other words, the study format.

Scenario	The predetermined sequence of events and actions required in the prototype.
Virtual environment	The levels and setting in which the scenario takes place.
VR system	A Head-Mounted Display (HMD) and its associated peripherals such as controllers and trackers
Simulator	The prototype developed during the thesis work including, but not limited to, the scenario and virtual environment.

Contents

1
Introduction

Virtual reality (VR) has become increasingly important in various fields and represents a significant advancement in interaction design and how users interact with digital content. Compared to traditional computer monitors, VR offers a more immersive experience that blurs the lines between the physical and virtual worlds. In recent years, VR technology has developed significantly and has become even more accessible and versatile due to the introduction of more affordable VR hardware on the market (Wohlgenannt et al., 2020). The evolution of technology has made it available to a wide range of users in fields such as education, healthcare, entertainment, and professional training. Although still in its early stages, VR has proven to be a valuable component in medical training due to its realistic simulation of complex medical scenarios without the risks of real-life training (Pottle, 2019). The platform offers a simulated and engaging learning environment for medical professionals to practice procedures and decision-making in a controlled setting. Additionally, VR provides more control over constructing scenarios within medical or military contexts that might otherwise be difficult or impossible to recreate practically.

The specific area of interest for the thesis is tactical combat casualty care (TCCC), mostly associated with medical and trauma care in the context of combat. Tactisim (2024) is a start-up company founded in 2022 on the idea of improving training in pre-hospital care using Augmented reality (AR) and VR simulators. They have since developed a simulator which immerses the learners in realistic critical scenarios such as addressing catastrophic bleeding. The simulator is designed to support the combination of virtual and real elements, e.g. a person wearing trackers being displayed for the user inside the virtual environment, giving the user a physical aspect to interact with. This caught the attention of the Swedish Home Guard for whom Tactisim are now developing a mobile application to help streamline their medical training (Tactisim, 2024). Based on the work of Tactisim, the thesis project will consist of developing a prototype of a casualty care simulator for standalone VR.

1.1 Research Question

The following research question will guide the research:

RQ: *What should be considered when designing a casualty care simulator for standalone VR systems?*

1.2 Aim

This thesis project will produce a prototype of a VR casualty care simulator for a standalone VR system designed to train users in handling medical emergency situations in an environment resembling the real-world contexts in which the knowledge will be used. The thesis aims to expand the understanding and practical application of VR technology in the context of casualty care simulators for standalone VR systems. The main contribution of the thesis is the VR prototype, but the thesis also aims to produce a set of design considerations for designing casualty care simulators for standalone VR systems.

2

Background

This chapter presents an overview of previous work and other pertinent literature on the subject of VR, casualty care, and provides illustrative examples of the current state of the technologies. In order to provide a comprehensive background for the thesis, an extensive literature and product review was conducted. A comprehensive search was conducted across a range of academic databases, including Google Scholar, PubMed, and IEEE Xplore, as well as academic journals pertinent to the fields of medical simulation, military training, and virtual reality technologies. Additionally, web search engines were utilised to identify similar products and current trends in the field.

2.1 Introduction to casualty care

For different kinds of casualty care, be it emergency or combat, there exists a plethora of simulators (Gillett et al., 2008). These simulators are intended to imitate real-life situations and aim to enable medical personnel to practice their skills in a controlled environment without endangering actual patients. Casualty care simulators are available in various forms, ranging from simple task trainers for practising specific procedures, such as wound suturing or airway management, to high-fidelity simulators that replicate the look, feel, and physiological responses of a human body (Gillett et al., 2008). These high-fidelity simulators may include features such as realistic anatomy, simulated bodily fluids, breathing mechanisms, and vital sign monitoring (Berkenstadt et al., 2003; Peckler et al., 2007). These features allow medical workers to practice a wide range of clinical skills, including patient assessment, diagnosis, treatment, and communication (Schulz et al., 2014).

Additionally, casualty care simulators in the military context are often part of TCCC training. TCCC is a set of military guidelines that emphasise rapid assessment, treatment, and evacuation of injured personnel in combat settings where medical resources may be limited (F. K. Butler & Kotwal, 2017). Simulated scenarios may include casualty evacuation procedures, combat triage, and managing multiple casualties simultaneously (F. Butler, 2017). Furthermore, casualty care simulators in the military may be used to train medical personnel in the use of specialised equipment and techniques specific to military medicine such as tourniquet application, combat dressing application, and the administration of battlefield medications (Scalese et al., 2022). Some specific use cases include treating gunshot wounds, blast injuries, burns, and other traumatic injuries resulting from explosions or combat situations (Berkenstadt et al., 2003; Pasquier et al., 2014).

2.2 Virtual Reality

VR systems have been explored and developed in military contexts since the 1960s (Fiani et al., 2020). The technology has developed drastically since then with more immersive software becoming widely available. In the 2010s, VR technology experienced significant advancements that transformed its application and made it more accessible to the general public (Wohlgenannt et al., 2020). This period saw the introduction of affordable VR systems, such as the Oculus Rift, HTC Vive, and later models like the Meta Quest which made VR more accessible to a broader audience (Wohlgenannt et al., 2020).

Moreover, the 2010s were marked by the development of VR software that could simulate complex and dynamic environments with a high degree of realism. These software advancements allowed for the creation of detailed and interactive scenarios (Marín-Morales et al., 2020). Since then, several studies have been conducted examining VR's viability for healthcare, military, fire department, and law enforcement sectors.

2.3 Systems and devices

Several casualty care simulators available today are based on physical simulation using either a human actor or a realistic doll. An example of this is Operative Experience which develops physical simulators (manikins) for training of medical procedures from casualty care to advanced surgeries (Operative Experience, 2024). Most relevant is their line of Tactical Casualty Care Simulators which includes interchangeable wounds, integrated software for monitoring vitals, and simulated scenarios and injuries (see figure 2.1). Furthermore, they are designed in such a way as to be able to be trained on in multiple different environments including outdoors in scenario-realistic environments (Operative Experience, 2024).

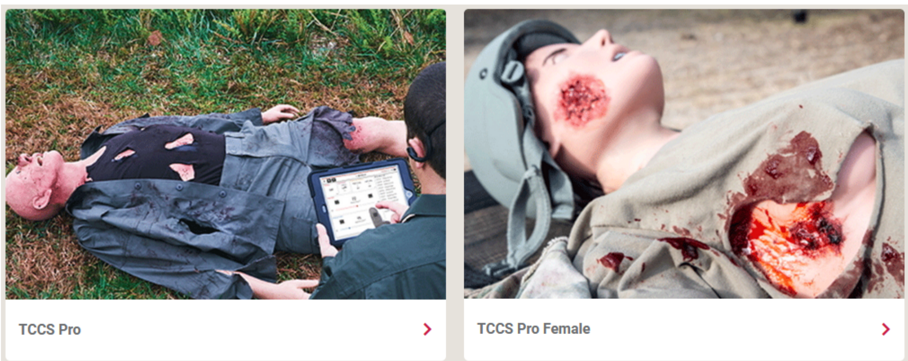


Figure 2.1: Two different casualty care simulators by Operative Experience (2024).

There have also been trials of simulators made with computer-generated graphics for casualty care in emergency departments. Already in the 2000s, Heinrichs et al. (2008) attempted to use an online game to train medical personnel to conduct surgery and practice teamwork in an emergency department context using neither human actors nor realistic manikins. Instead, the physical manikins were replaced by computer-generated patients on which the team members performed surgery as depicted in figure 2.2. While the study did not use any head-mounted display (HMD) technology to achieve an elevated level of immersion, the participants generally thought the virtual simulation was effective for practising clinical skills and cooperation between team members (Heinrichs et al., 2008).

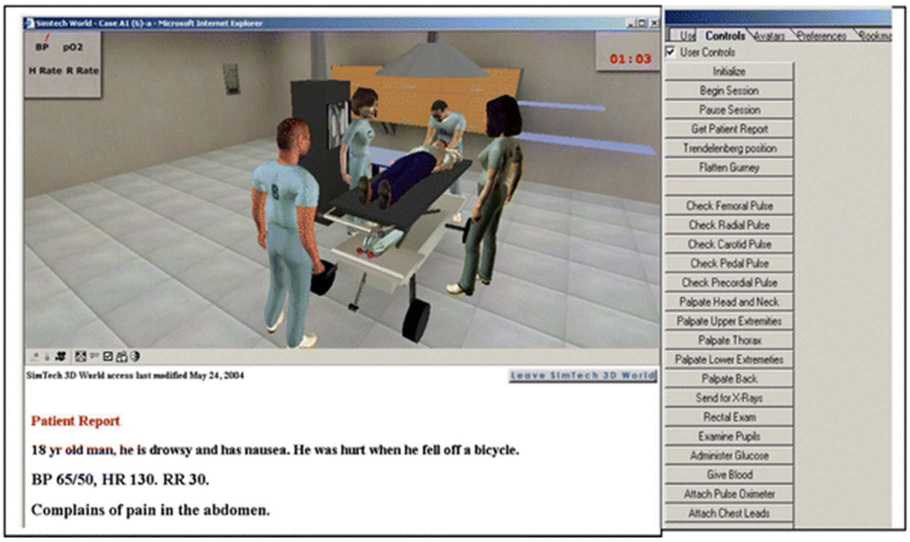


Figure 2.2: Virtual surgery simulator from the 2000s (Heinrichs et al., 2008)

Street Smarts VR (2021) is a company that specialises in VR solutions tailored mainly to police and military training. They offer a range of products and scenarios catered to various law enforcement and defence agencies that focus on simulating realistic scenarios encountered in these professions. Utilising the HTC Vive Pro for hardware, the simulator is adaptable and employs different props as controllers to match the specific simulated scenario. For instance, in a military training exercise, an assault rifle might serve as the controller, whereas a baton would be used in police training scenarios. The scenarios are designed to have branching narratives to allow for multiple possible outcomes to every scenario. This dynamic structure allows trainees to experience a wide range of situations, from routine operations to high-stress crisis events, thereby preparing them for a variety of real-life encounters. After each scenario is completed, the user is shown a debrief with quantitative data that has been measured throughout the scenario.

Lovreglio et al. (2021) conducted a study comparing VR and video-based training for learning how to operate a fire extinguisher according to the Pull, Aim, Squeeze and Sweep (PASS) procedure. The video-based training consisted of watching a video going through each step in order using animations, illustrating how to perform the action. The VR-based training consisted of acting out a scenario of a fire in a warehouse where the participant had to locate and acquire the fire extinguisher and then follow the PASS procedure to put out the fire. The equipment used to present the virtual environment was the HTC Vive. The results of the study, conducted with 93 participants, showed that training using the VR simulator proved more effective in knowledge acquisition and retention, self-efficacy, and gave a higher perception of recommendations efficacy and simplicity. The study also showed that participants from the group training on VR maintained their level of self-efficacy more than three weeks after the training whilst participants from the video-based training group had significantly decreased in self-efficacy. The feedback from the participants revealed a need for improved content, difficulty, and realism. For example, deciding between multiple different extinguishers depending on the source and size of the fire.

Barteit et al. (2021) conducted a systematic review of 27 studies with a global health perspective of low- and middle-income countries (LMICs) to examine the effectiveness of Extended reality (XR) HMDs for medical education. All studies were conducted with healthcare professionals and satisfied the criterion of comparing and evaluating the effectiveness of an HMD intervention and another method of learning. Out of the 27 studies, 17 of them found that solutions using HMDs were effective, and another four found that they could be effective for certain aspects of learning and training in the field of medicine. Barteit et al. (2021) do, however, express that many of the studies were of small scale and urge future research to "rigorously evaluate XR-based HMDs with AR and VR implementations, particularly in LMICs, to better understand the strengths and shortcomings of HMDs for medical education" (Barteit et al., 2021, p.2).

Mehrfard et al. (2021) conducted a study with the goal of identifying VR technologies suitable for clinical education. Specifically, Mehrfard et al. (2021) evaluated the technologies in isolation, and not in a specific context such as learning how to perform a medical procedure. Mehrfard et al. (2021) evaluated 10 different VR technologies based on specifications and metrics such as resolution, weight, and torque on the neck. The three best technologies, namely Samsung Odyssey+, HTC Vive Pro, and Oculus Rift S, were then further evaluated in a user study with 27 participants based on text readability, comfort, and contrast perception. Mehrfard et al. (2021) found the HTC Vive Pro to be the best overall platform, being notably outperformed in only readability by the Oculus Rift S.

Stress is an important aspect that needs to be considered in a casualty care simulator. Binsch et al. (2021) claim that implementing stress inducement is crucial because it enables the user to learn how to better handle the stress that may arise in the real-life scenarios that the simulator emulates. Binsch et al. (2021) conducted a study with 66 cadets from the Dutch Army who were put into a military convoy simulation. During the simulations, the participants were posed multiple-choice questions related to their surroundings such as present people, buildings, and road signs. The simulation contained four phases which were designed to become increasingly stressful to the participants. Physiological metrics such as heart rate and blood pressure were monitored and the participants were also asked to fill out a questionnaire after finishing the scenario. Interestingly, the physiological stress in the participants seemed to lessen gradually throughout the scenario even though the scenarios were designed to become increasingly stressful. Binsch et al. (2021) theorised that it could be attributed to the novelty of the scenario being the primary effective stress stimulus rather than the design of the scenarios themselves. Thus, they conclude that to maintain a reasonable level of stress in VR simulator training, “scenario designers should look to increase task difficulty or complexity, and implement consequences of performance” (Binsch et al., 2021, p.192).

Another study that explored stress as one of the relevant variables was conducted on Dutch police officers. Kleygrewe et al. (2023) investigated the effects of VR scenario-based training (SBT) on police officers’ training responses compared to real-life scenario-based training (RL SBT). The scenario used for the RL SBT was a large-scale scenario with several actors while the VR SBT used three smaller scenarios. The study measured a few relevant variables like heart rate, mental effort, and perceived stress. It found that while the maximum heart rate and the average level of physical activity were significantly higher in RL SBT, the mental effort was significantly higher in VR SBT. No meaningful differences were found in average heart rate and perceived stress. Based on the findings of the study, Kleygrewe et al. (2023) suggest VR as a promising complement to police training with scenarios demanding significant mental rather than physical exertion. Kleygrewe et al. (2023) further found that invested mental effort in VR SBT was a bit higher for participants without previous VR knowledge or experience, but they found no difference in perceived stress. Kleygrewe et al. (2023) theorise that this variance in mental effort observed during the study is related to the disparity in VR familiarity among participants. Those with limited exposure to VR environments likely required more cognitive resources to navigate and interpret the virtual setting in comparison to participants with more VR experience.

2.4 Research Problem

Tactisim is mainly interested in three platforms for their simulator: high-end VR systems such as the Varjo XR-3, standalone VR systems such as the Meta Quest 3, and mobile platforms. At this point, Tactisim has a prototype of their casualty care simulator which can be run on a high-end VR system. The limitations of high-end VR restricts the simulator in several areas:

- **User movement**: The tethered HMD restricts the user’s ability to move which in turn impairs the user experience.
- **Awareness**: The tethered HMD requires the user to always be mindful of the cable while using the simulator which impairs awareness.
- **Accessibility**: A high-end VR system, a room with tracking cameras, and a computer with enough performance to run the simulator are costly investments.

- **Physical space**: The physical space required for a VR setup with tracking cameras and a computer is much larger compared to standalone VR systems and mobile.

- **Troubleshooting**: A setup with multiple components means that more things can malfunction and troubleshooting thus becomes harder.

Furthermore, Tactisim is developing a version of the casualty care simulator for the mobile platform. The mobile platform will have significant differences compared to the VR version of the simulator because of the inherent limitations and characteristics of the mobile platform such as:

- **User input**: Mobile devices are built to take physical user input solely from the touchscreen which limits interaction capabilities.

- **Presence**: The mobile platform lacks the level of presence of VR which limits its capabilities to emulate the high stakes that real-world scenarios may include, such as psychological and physiological stress.

- **Distractions**: Mobile devices are multi-purpose, meaning that notifications from other apps can distract users during training sessions which reduces the effectiveness of the simulation.

Standalone VR systems such as the Meta Quest 3 hypothetically present a compelling compromise between the mobile platform and the high-end VR platform which is why it is the chosen platform for the prototype in the thesis.

3

Theory

This chapter delves into a few theoretical frameworks that are planned to be used during the thesis. These theories offer a few perspectives for understanding user interaction, acceptance, and cognitive processes in technology use and design which will help guide the development of a viable and user-centred solution. Table 3.1 provides a short overview of the theories in the proceeding sections, explaining why and how they will be used throughout the thesis project.

Table 3.1: A summary of why the theories were chosen and their usage in the thesis

Theory	**Why?**	**Usage**
Activity Theory	To provide a holistic framework to study user interactions and motivations	Understanding the context of user learning and motivations
Technology Acceptance Model	To understand key factors that influence users' perception of the prototype's usability	Starting point for what to evaluate in the study
Cognitive Load Theory	To understand how to balance the different types of cognitive load to foster an optimal learning environment	Guide the prototyping and be a starting point for what to evaluate in the study
Design Thinking	To establish a structure in the design process used during prototyping and evaluation	Prototyping
Research through Design	To produce both academic and practical insights through iterative development and testing of the prototype	Prototyping

3.1 Activity Theory

Activity theory (AT) is a comprehensive framework for understanding human behaviour and consciousness within an environment (Karanasios et al., 2021). It conceptualises human activities as systemic and structured interactions between subjects (actors) and their environment and states that subjects are driven by objects (objectives or goals) that are mediated by tools and artefacts. According to (Karanasios et al., 2021), tools refer to physical and symbolic mediators such as language and technology that facilitate and shape human activity. Artefacts are seen as broader mediating entities in human activities. Artefacts encompass tools but also the accumulated underlying knowledge or cultural heritage that inform and shape activities such as cultural and societal norms which the theory states always impact how subjects act in an environment.

AT distinguishes between the concepts of action and operation. McAvinia (2016) describes actions as the conscious and object-directed processes that subjects undergo to achieve a goal. Actions are mediated by tools and are characterised by the intent of the subject. In other words., actions are always performed with a goal in mind. In contrast, McAvinia (2016) describes operations as the automatic processes that support actions. Operations are typically not conscious and do not have objects in themselves but depend on the situation or the conditions in which the action is performed. This means that operations are the routines or procedures that become automated in subjects through practice and do not require conscious thought once mastered. For example, an operation might be the specific hand movements and force application needed to perform CPR effectively. Initially, a learner might concentrate heavily on these movements, but as they become more skilled, the movements become automatic operations that support the broader action of performing CPR.

An essential aspect of AT is the dynamic relationship between actions and operations. Actions can become operations as they are internalised and automated through practice, and what was once a conscious effort becomes an unconscious process. Conversely, operations can become actions again if a problem arises or if the conditions change which requires the subject to consciously reconsider and adapt their approach (Kaptelinin, 2014; McAvinia, 2016; Nardi, 1995).

When designing systems, AT provides a holistic approach that goes beyond the interaction between the user and technology to include the broader context of use. This involves a deep understanding of the objects that subjects aim to achieve, the tools and artefacts that are available for achieving these goals, and the community within which the subjects act. The community includes things such as norms, rules, and roles that guide the behaviours of the subject (Karanasios et al., 2021).

In this thesis, AT will be used to understand how users learn and how their individual contexts and perceptions of the scenario affect their interactions and decision-making.

3.2 Technology Acceptance Model

The Technology Acceptance Model (TAM) is a widely used theoretical framework in the field of information systems and technology adoption research. Developed by Davis (1985), TAM seeks to understand and predict users' acceptance of new technology by examining their attitudes and perceptions. The theory suggests two primary factors that influence individuals' intention to use a particular technology:

- **Perceived Usefulness (PU)**: Davis (1985) defines perceived usefulness as the degree to which a person believes that using a specific technology will enhance their performance or productivity in accomplishing tasks. In the context of TAM, users assess the utility and benefits they expect to derive from adopting the technology. If users perceive a technology as useful, they are more likely to accept and use it.

- **Perceived Ease of Use (PEOU)**: Davis (1985) defines perceived ease of use as the degree to which a person believes that using a particular technology will be free from effort or cognitive strain. It encompasses factors such as the simplicity of the interface, ease of navigation, and the amount of learning required to operate the technology effectively. A technology that is perceived as easy to use is more likely to be accepted by users as it reduces barriers to adoption and encourages engagement.

PU and PEOU are key determinants of users' attitudes and intentions toward adopting a new technology. When users perceive a technology as both useful and easy to use, they are more likely to exhibit positive attitudes towards it and express a greater intention to use it in the future (Davis, 1985). TAM has been applied across various domains including healthcare, education, business, and consumer technology to understand and predict user acceptance of new technological innovations (Aggelidis & Chatzoglou, 2009; Wang et al., 2008).

In this thesis, TAM will be used to guide the decisions on which factors to consider in the study and in turn affect both the format of the study and which methods to use for gathering data.

3.3 Cognitive Load Theory

Cognitive Load Theory (CLT) is a theoretical framework that explores how the human cognitive system processes information and more specifically how cognitive load affects learning, problem-solving, and knowledge retention. In the article credited with coining the theory, Sweller (1988) suggests that there are limits to the amount of information that individuals can process in their working memory at any given time. When the cognitive load exceeds these limits, learning becomes less effective and performance may suffer. In later works, Sweller (2010) identifies three types of cognitive load:

- **Intrinsic Cognitive Load**: Intrinsic cognitive load refers to the inherent complexity of the learning materials or tasks. Some tasks naturally require more mental effort and cognitive resources to understand and perform. For example, complex medical procedures or casualty care scenarios inherently impose a high intrinsic cognitive load on learners.

- **Extraneous Cognitive Load**: Extraneous cognitive load is caused by factors external to the learning materials or tasks, such as poorly designed instructional materials or distracting environmental conditions. Extraneous cognitive load detracts from learning by placing unnecessary demands on working memory. Effective learning material therefore helps to minimise extraneous cognitive load and optimises learning outcomes.
- **Germane Cognitive Load**: Germane cognitive load refers to the cognitive effort devoted to schema acquisition and automation. When learners successfully process and organise information into meaningful mental structures, so-called schemas, they can reduce cognitive load by automating routine cognitive processes. Germane cognitive load is beneficial for learning because it facilitates the construction of long-term memory and promotes expertise development.

Achieving a balance between the three types of cognitive load is crucial for instructional design and the development of educational materials. An effective way to manage intrinsic cognitive load is by breaking down complex information into smaller, more manageable chunks. This approach, known as ”chunking”, helps learners to process and understand the material without overwhelming their working memory (Kahn et al., 1990). Additionally, minimising extraneous cognitive load is essential and can be accomplished by eliminating unnecessary information and distractions, simplifying instructional design, and utilising clear and concise language. Finally, enhancing germane cognitive load involves creating learning activities that encourage deep processing and facilitate the integration of new information into existing cognitive schemas. Effective instructional design that takes into account cognitive load theory usually leads to more efficient and meaningful learning experiences for learners (Merriënboer & Sweller, 2005).

In this thesis, CLT will be used to guide the decisions on which factors to consider in the study and in turn affect both the format of the study and which methods to use for gathering data. CLT will also help keep cognitive load in mind when developing the prototype.

3.4 Design Thinking

Design Thinking (DT) represents a user-centred approach to problem-solving that integrates the needs of people, the possibilities of technology, and the requirements for business success (Luchs et al., 2015). This method has gained prominence within various fields, including Human-Computer Interaction (HCI), for its efficacy in addressing complex problems through a systematic yet flexible process. In the context of academic research, the exploration of DT not only highlights its practical applications but also underscores its theoretical underpinnings and methodological structure (Luchs et al., 2015). Luchs et al. (2015) identify a two-phase approach to DT: Identify and Solve, each comprising two interchanging modes, thereby offering a comprehensive overview of this iterative process.

The Identify phase is foundational to the DT process and aims to understand the problem space with depth and breadth (Luchs et al., 2015). This phase is realised through two critical modes: *Discover* and *Define*. The *Discover* mode is characterised by an open-minded exploration of the problem domain. It involves engagement with the target users or stakeholders to gain insights into their experiences, needs, and challenges. Techniques such as observations, interviews, and immersive experiences are used to gather qualitative data. This exploratory mode encourages designers to look beyond the obvious, challenge assumptions, and discover latent needs that users themselves might not be aware of (Luchs et al., 2015).

Building on the insights gathered during the Discover mode, the *Define* mode seeks to synthesise this information into a clear and actionable problem statement (Luchs et al., 2015). This mode involves critical analysis and filtering of the collected data to identify patterns, themes, and key insights. The outcome is a well-articulated definition of the problem, often represented as a point of view statement that encapsulates the users' needs, the context of the problem, and the potential for impact. This definition sets a focused direction for ideation and solution development (Luchs et al., 2015).

The Solve phase transitions from problem identification to the generation and implementation of solutions (Luchs et al., 2015). It is put into practice through the *Create* and *Evaluate* modes. In the *Create* mode, the emphasis shifts to ideation and the generation of innovative solutions. Leveraging the clearly defined problem statement, designers brainstorm a wide range of ideas without constraints, encouraging creativity and out-of-the-box thinking. This mode often involves collaborative workshops, sketching, prototyping, and other methods to visualise and concretise ideas. The goal is to explore a broad spectrum of potential solutions before converging on the most promising ones for further development (Luchs et al., 2015).

The *Evaluate* mode is critical for refining solutions through iterative testing and feedback (Luchs et al., 2015). Prototypes, ranging from low to high fidelity, are subjected to user testing to gather insights into their usability, effectiveness, and user satisfaction. This mode involves observing users interacting with the prototypes, collecting feedback, and identifying areas for improvement. The iterative nature of the Evaluate mode ensures that solutions are continuously refined based on real user feedback and leads to more user-centred and practically viable outcomes (Luchs et al., 2015).

Luchs et al. (2015) further stress that the two phases and their interchanging modes should not be treated as a linear process even though they are usually introduced as that. Instead, they should be seen as parts of an iterative procedure that helps designers produce potential solutions despite incomplete knowledge of the task at hand. These preliminary solutions should then serve as a basis for deeper and improved understanding, insights, and solutions. The use of the term "mode" instead of "step" reflects this flexible and iterative process and emphasises that the progression through these modes is not fixed but varies depending on the project needs, objectives, constraints, and the team's judgment on the progress and direction of the project (Luchs et al., 2015).

In this thesis, DT will be used to guide the prototyping and design process, providing structure.

3.5 Research through Design

Research through Design (RtD) is a methodology in the fields of design and human-computer interaction. This approach is distinct in that it employs the design process as a tool for generating new knowledge and insights. The core of RtD lies in its iterative process of designing, creating, and reflecting which allows researchers and designers to engage with both the material and conceptual aspects of their inquiries (Zimmerman et al., 2007). RtD fundamentally acknowledges the notion that the practice of design itself is not only a creative endeavour but also a form of inquiry. In other words, it is a methodological approach that actively engages with and interrogates the material and conceptual realms to generate deep and insightful knowledge (Zimmerman et al., 2010). Designers engage with materials, technologies, and user contexts through the act of making and thereby uncover insights that otherwise would not be accessible through observation or analysis alone. This hands-on engagement facilitates a deeper understanding of the problem space and enables researchers to identify and articulate needs, opportunities, and design implications that might otherwise remain hidden (Zimmerman et al., 2010).

The RtD process typically involves several key stages: identifying an area of interest or a research question, developing prototypes or design artefacts, deploying these artefacts in real or simulated contexts, and then reflecting on the outcomes of these interventions (Zimmerman et al., 2007). This reflection is not only subjective but is also grounded in the theoretical frameworks and research questions that motivated the project. The insights gained through this reflective process contribute not only to the specific project at hand but also to the broader knowledge base of the field. One of the strengths of RtD is its capacity for generating experiential knowledge that comes from direct experience with designed artefacts and systems. This is particularly valuable in understanding human-centred design issues where the nuances of human interaction with technology are knotty and many-sided. RtD allows researchers to explore these interactions in-depth and gives insights into user behaviours, preferences, and experiences that would be difficult to capture through more conventional research methods (Zimmerman et al., 2007).

In this thesis, RtD will be used to guide and give structure to the prototyping and design process.

4 Methodology

This chapter is split up into sections that correspond to the stages of the design process. Section 4.2 diverges slightly from the other sections to provide a higher-level abstraction, focusing on the methods for conducting studies. In other words, it provides the context and reference frame for how the methods mentioned in sections 4.4 and 4.5 can be used.

4.1 Prototyping

There exists a smorgasbord of prototyping methods in the world of interaction design. On an elevated level, the methods are divided into two categories: Low-fidelity prototypes and high-fidelity prototypes. Each category is used to help a concept grow from an idea to a finished product and has its own strengths and weaknesses.

4.1.1 Low-fidelity prototyping

Low-fidelity prototypes are effective at the beginning of the design process when the concept is not yet developed. It uses rudimentary tools and materials for ideation and iteration such as sketches, paper, or digital wireframes to help articulate and validate design concepts (Sharp et al., 2019, Chapter 12). Despite their simplicity, low-fidelity prototypes generally excel in creating clear communication among stakeholders and provide a platform for early-stage feedback and iteration without caring too much about the minute details of the concept (Benyon, 2010, Chapter 8). However, the simplicity of low-fidelity prototypes generally limits the ability to accurately simulate user interactions and understand the nuances of the final product (Sharp et al., 2019, Chapter 12).

Storyboarding is a low-fidelity prototyping technique that is often employed to depict scenarios to illustrate how users interact with a prototype (Sharp et al., 2019, Chapter 12). Storyboarding consists of creating a series of sketches to explore and visualise a potential user's journey through a task or activity and usually incorporates the concept as a driving force which allows stakeholders to gain insights into how the concept may be used as a developed product (Sharp et al., 2019, Chapter 12).

Wizard of Oz is another low-fidelity prototyping method which makes use of a human operator to simulate the functionality of the prototype from behind the scenes (Sharp et al., 2019, Chapter 12). This allows designers to examine how users would interact with the product or system without having to implement its functionality.

Paper prototyping is yet another low-fidelity prototyping method which is good for creating quick and easy visualisations which clearly unfinished nature invites criticism and feedback by decreasing the likelihood of users holding back due to fear of hurting the researchers' feelings (Dam & Teo, 2024).

4.1.2 High-fidelity prototyping

High-fidelity prototypes embody a more refined level of design to emulate the look, feel, and functionality of the intended product. They usually use more advanced design tools and technologies such as interactive mockups, higher-quality building materials, or coded simulations and thus provide a holistic representation of the envisioned user experience (Sharp et al., 2019, Chapter 12). Their high attention to detail also allows for comprehensive user testing which enables designers to gain valuable insights into user behaviour and interaction patterns.

4.2 Study Format

Usability testing is a research method used to evaluate how easy and user-friendly a product or system is by observing real users as they attempt to complete a task using the product or system (Sharp et al., 2019, Chapter 14). By observing the users interact with the system and following up with methods such as interviews and questionnaires (see section 4.4) both quantitative data and qualitative data can be gathered and used to evaluate the system. This method can be used in combination with the following methods of experimental design to facilitate the comparison between different systems or user groups.

Between-subjects design involves assigning groups of participants which only test one condition of the independent variable (IV) (MacKenzie, 2013, Chapter 6). In other words, if the study aims to compare two different systems, each group would only test one of the systems.

Within-subjects design instead lets all participants test all conditions of the IV (MacKenzie, 2013, Chapter 6). In the same example as above, all users would test both systems.

4.3 Sampling

This section contains several methods for gathering participants for studies.

Probability sampling, also called random sampling, is a collection of sampling methods in which every member of the population has a known and equal chance of being selected (Government of Canada, 2021). This method aims to ensure that the sample accurately represents the population and allows for generalisable and statistically significant results. However, probability sampling has several drawbacks, chiefly relating to its inherent complexity. Namely, probability sampling methods often require detailed knowledge of the entire population and a comprehensive list of all members which can be difficult, expensive, and time-consuming to compile (Government of Canada, 2021).

Simple random sampling is the most straightforward probability sampling method. In this approach, each member of the population is assigned a unique identifier, and a random selection process is used to choose the sample (Thompson, 2012). This can be done using random number tables, computer algorithms, or drawing names from a hat (Thompson, 2012). The key advantage of simple random sampling is its simplicity and the equal probability of selection for each population member, ensuring unbiased representation. However, it can be impractical for large populations where listing every individual is challenging (Government of Canada, 2021).

Systematic sampling involves selecting subjects from an ordered list at regular intervals (Government of Canada, 2021). For instance, if a researcher wants to sample 100 individuals of 1000, they might choose every 10th person on the list. This method is easy to implement and ensures even coverage of the population (Thompson, 2012). However, systematic sampling can introduce bias should there be a hidden pattern in the population list that corresponds to the sampling interval (Government of Canada, 2021).

Stratified sampling aims to improve representativeness by dividing the population into subgroups, or strata, based on specific characteristics such as age, gender, or income level (Thompson, 2012). A random sample is then drawn from each stratum in proportion to its size in the population. This method ensures that all relevant subgroups are adequately represented in the sample which enhances the results' accuracy and reliability (Thompson, 2012). While stratified sampling requires detailed knowledge of the population characteristics and can be more complex to administer, it is particularly useful in heterogeneous populations where certain subgroups might otherwise be underrepresented (Government of Canada, 2021).

Non-probability sampling methods are often used in research when it is impractical or impossible to conduct probability sampling (Baker et al., 2013). Non-probability sampling does not involve random selection which means that some members of the population have no chance of being selected, and the probability of selection for those chosen cannot be calculated (Baker et al., 2013). Among the most common non-probability sampling methods are convenience sampling and snowball sampling, each with distinct processes, recruitment strategies, and potential biases.

Convenience sampling involves selecting subjects that are easiest for the researcher to access, chiefly individuals who are readily available and willing to participate (Marshall, 1996). Recruitment can happen in various settings such as shopping malls, universities, workplaces, or online platforms. This method is notably prone to selection bias since the sample is not randomly selected and could lead to a bias towards certain participant characteristics (Baker et al., 2013). Additionally, the results of convenience sampling cannot always be generalised to a broader population because the sample may not be representative of it (Marshall, 1996).

Consecutive sampling involves selecting every subject who meets the inclusion criteria and is available during a specific period of time or until a particular sample size is reached (Broeck & Brestoff, 2013). The process of consecutive sampling begins by establishing clear inclusion and exclusion criteria for the study. Once these criteria are set, researchers recruit subjects consecutively as they become available. However, consecutive sampling is not without its biases. One primary concern is the potential for temporal bias, as the sample may be influenced by time-specific factors (Schuster & Powers, 2005).

Purposive sampling, also known as judgemental sampling, involves selecting subjects based on specific characteristics or criteria set by the researcher and seeks to select individuals who are most likely to provide relevant and rich information (Marshall, 1996). Subjects are chosen based on the researcher's judgment such as selecting individuals who have undergone specific training for a study on its impact. This method can introduce researcher bias as the selection process is subjective and influenced by the researcher's own biases and assumptions (Nikolopoulou, 2022a).

Snowball sampling begins with a small group of initial test subjects who meet the criteria for the study. These initial subjects then recruit future subjects from among their acquaintances and the future subjects may continue the process with each recruit referring more subjects (Naderifar et al., 2017). This method, like convenience sampling, also carries biases. For example, the sample generated through snowball sampling is likely to be biased towards individuals more connected to the initial subjects which could lead to homogeneity in the sample and a limited diversity of perspectives (Emerson, 2015).

Quota sampling involves dividing the population into subgroups, so-called quotas, and selecting a predetermined number of subjects from each subgroup to ensure the representation of certain characteristics in the sample (Yang & Banamah, 2014). It is essentially the non-probabilistic version of the previously mentioned stratified sampling. To make quota sampling effective, the quotas should be based on mutually exclusive features that make the placement of subjects into the different subgroups straightforward (Nikolopoulou, 2022b). The quotas can be based on several characteristics such as age, gender, income level, and educational background, and subjects are recruited to fill these quotas, often in combination with other sampling methods (Yang & Banamah, 2014). Since quota sampling usually requires using the previously mentioned sampling methods, it is subject to the same biases that those methods have such as selection bias (Nikolopoulou, 2022b).

4.4 Data gathering

This section contains various methods for gathering both quantitative and qualitative data.

4.4.1 Observation

Observation can be divided into two categories: direct and indirect observation. *Direct observation*, specifically in controlled environments, entails studying and observing participants performing specified tasks in a controlled time and place (Sharp et al., 2019, Chapter 8). This can be used in combination with video recording to help capture the participants' behaviour from facial expressions to body language (Sharp et al., 2019, Chapter 8). An important aspect when conducting studies in a controlled environment is to prepare a script in advance, covering the process of greeting the participants, informing them of how and what they are going to participate in, and clarifying their rights (Sharp et al., 2019, Chapter 8). This way, the researchers can ensure that all participants are treated in the same manner and that they adhere to the principle of informed consent (University of Oxford, 2021). A commonly used technique applied when in this type of environment is the *Think-Aloud* technique which consists of the user explaining their thought process by narrating what they are doing and importantly why they are doing it (Sharp et al., 2019, Chapter 8).

Direct observation can also be conducted in the field. This would be applicable when the object of study is something naturally occurring and not something which has to be carefully set up manually and tested in a dedicated space and time (Sharp et al., 2019, Chapter 8). An important aspect to consider when conducting this type of observation is to have a framework for what form any eventual observations will take. That is what information about the observation should be recorded. The complexity of the framework has to be adjusted to the context and purpose of the study. A framework can be as simple as recording who, where, and what (Sharp et al., 2019, Chapter 8).

Indirect observation is used when the form of study does not allow for direct observation. This could be due to the duration of the study spanning over a long period of time where observers cannot always be present (Sharp et al., 2019, Chapter 8). Diaries and interaction logs are common ways of accomplishing this. Diaries are focused on the user recording their activities and their reflections on them whilst interaction logs use software to log user activity such as key presses or time spent on a website (Sharp et al., 2019, Chapter 8).

4.4.2 Interviews

Unstructured interviews are meant for collecting qualitative data and consist of asking open-ended questions in order to allow for diverse answers (Sharp et al., 2019, Chapter 8). By asking questions without an expected answer format, the interviewee is free to answer in any way they like. This is meant to produce rich data, providing deep insight into the topic at hand and allowing for the interviewee to bring up points not considered by the interviewer (Sharp et al., 2019, Chapter 8). When conducting unstructured interviews, the interviewer must find a balance between getting answers to specific questions through probing and letting the conversation flow in order to explore new or unexpected territories (Sharp et al., 2019, Chapter

8).

Structured interviews involve asking predetermined questions, similar to those in a questionnaire, to each participant in a standardised manner (Sharp et al., 2019, Chapter 8). These questions are typically closed and brief which means that the participant only can choose from a set of predetermined options. As a result, the wording and order of the questions in a structured interview are always the same (Sharp et al., 2019, Chapter 8). The use of solely closed questions renders structured interviews best for situations where the goals are clear and time per participant is limited (Sharp et al., 2019, Chapter 8).

Semi-structured interviews include both open and closed questions. Specifically, the interviewer starts by asking planned questions and follows them up with open-ended probing questions to get the interviewee to elaborate on their answer (Sharp et al., 2019, Chapter 8). The follow-up questions need to be formulated and asked with careful consideration. It is important that they not be suggestive or based on an assumption made by the interviewer instead of something explicitly expressed by the interviewee (Sharp et al., 2019, Chapter 8).

The previously mentioned interview methods are conducted using one interviewer and one interviewee, but sometimes interviews may be conducted with a group of interviewees. An example method of conducting interviews like this is *focus groups*. Involving more than one interviewee, focus groups allow for diverse perspectives based on shared rather than individual experiences and are valuable for understanding conflicting expectations among participants (Sharp et al., 2019, Chapter 8). Participants are selected to represent the target population and discussions are guided by a preset agenda with the flexibility to explore topics as they are brought up (Sharp et al., 2019, Chapter 8).

All interview methods can be used in combination with audio recording. This can be used to transcribe the interview and facilitate the process of analysing the interview. However, the use of audio recordings may not be as necessary for structured interviews because of the static and closed nature of the questions and answers. Instead, it might simply be more time-efficient to note the responses down during the interview, effectively eliminating the need to go back and listen afterwards.

4.4.3 Questionnaires

Questionnaires can include both open and closed questions and are often used to gather demographic data (Sharp et al., 2019, Chapter 8). The benefit of using a questionnaire is that it can be distributed to a large number of people and can in contrast to interviews be completed remotely and on the participants' own time (Sharp et al., 2019, Chapter 8). That does mean, however, that significant effort needs to be put into formulating the questions since there is no interviewer present to help clarify if the question is misunderstood by the participant. Asking specific questions with a wide range of answer options, including "none of the above", is important to address this issue (Sharp et al., 2019, Chapter 8).

A specific subset of questionnaires is *Standardised questionnaires* which are typically very closed and rigid. They are often used as measurement scales to allow for approximate, standardised scoring of a certain aspect.

System Usability Scale (SUS), developed by John Brooke (1996) is a standardised ten-item scale with the purpose of measuring the usability of a system. The ten items are:

1. I think that I would like to use this system frequently
2. I found the system unnecessarily complex
3. I thought the system was easy to use
4. I think that I would need the support of a technical person to be able to use this system
5. I found the various functions in this system were well integrated
6. I thought there was too much inconsistency in this system
7. I would imagine that most people would learn to use this system very quickly
8. I found the system very cumbersome to use
9. I felt very confident using the system
10. I needed to learn a lot of things before I could get going with this system

Each item is measured on a 5-point Likert scale. The positively formulated items (i.e. all odd-numbered items) are scored by subtracting 1 from the scale position and all the negatively formulated items (i.e. all even-numbered items) are scored by subtracting the scale position from the maximum value, five that is. According to Soegaard (2023), a SUS score above 70 is considered good and a score above 85 excellent.

NASA Task Load Index (NASA-TLX) is a tool for measuring a user's perceived workload when performing a task (Hart, 2006). The user rates their experience across six different categories:

- Mental Demand - How mentally demanding was the task?
- Physical Demand - How physically demanding was the task?
- Temporal Demand - How hurried or rushed was the pace of the task?
- Performance - How successful were you in accomplishing what you were asked to do?
- Effort - How hard did you have to work to accomplish your level of performance?

- Frustration - How insecure, discouraged, irritated, stressed, and annoyed were you?

The scores are mapped to range from zero to 100 and are then weighted based on the user's personal definition of workload through a series of pairwise comparisons between the subscales. According to Hertzum (2021), analysing TLX scores from 556 studies, mean TLX scores and standard deviation for the domains of emergency response and engineering design were 46±11 and 38±12 respectively.

Raw TLX (RTLX) is a variant of the standard NASA-TLX which skips the step of assigning weights to the subscales and instead calculates a single score by either averaging or adding together the scores from all six subscales (Hart, 2006). The RTLX has become increasingly popular because it is easier to apply (Hart, 2006).

The Witmer and Singer (1998) *Presence Questionnaire* (WS) is a standardised questionnaire consisting of 32 questions used to measure a user's perceived presence in a virtual environment. The 32 questions are:

1. How much were you able to control events?
2. How responsive was the environment to actions that you initiated (or performed)?
3. How natural did your interactions with the environment seem?
4. How completely were all of your senses engaged?
5. How much did the visual aspects of the environment involve you?
6. How much did the auditory aspects of the environment involve you?
7. How natural was the mechanism which controlled movement through the environment?
8. How aware were you of events occurring in the real world around you?
9. How aware were you of your display and control devices?
10. How compelling was your sense of objects moving through space?
11. How inconsistent or disconnected was the information coming from your various senses?
12. How much did your experiences in the virtual environment seem consistent with your real-world experiences?
13. Were you able to anticipate what would happen next in response to the actions that you performed?
14. How completely were you able to actively survey or search the environment using vision?

15. How well could you identify sounds?
16. How well could you localize sounds?
17. How well could you actively survey or search the virtual environment using touch?
18. How compelling was your sense of moving around inside the virtual environment?
19. How closely were you able to examine objects?
20. How well could you examine objects from multiple viewpoints?
21. How well could you move or manipulate objects in the virtual environment?
22. To what degree did you feel confused or disoriented at the beginning of breaks or at the end of the experimental session?
23. How involved were you in the virtual environment experience?
24. How distracting was the control mechanism?
25. How much delay did you experience between your actions and expected outcomes?
26. How quickly did you adjust to the virtual environment experience?
27. How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?
28. How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?
29. How much did the control devices interfere with the performance of assigned tasks or with other activities?
30. How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?
31. Did you learn new techniques that enabled you to improve your performance?
32. Were you involved in the experimental task to the extent that you lost track of time?

These questions relate to different categories of factors which contribute to a sense of presence: Control factors such as degree of control, Sensory factors such as degree of movement perception, Distraction factors such as interface awareness, and Realism factors such as information being consistent with the objective world (Witmer & Singer, 1998).

Slater-Usoh-Steed Questionnaire (SUS-Q) is another standardised questionnaire aimed at measuring presence in a virtual environment (Usoh et al., 2000). It consists of six items measured on a 7-point Likert scale. The items are:

1. Please rate your sense of being in the virtual environment, on a scale of 1 to 7, where 7 represents your normal experience of being in a place.

2. To what extent were there times during the experience when the virtual environment was the reality for you?

3. When you think back to the experience, do you think of the virtual environment more as images that you saw or more as somewhere that you visited?

4. During the time of the experience, which was the strongest on the whole, your sense of being in the virtual environment or of being elsewhere?

5. Consider your memory of being in the virtual environment. How similar in terms of the structure of the memory is this to the structure of the memory of other places you have been today? By 'structure of the memory' consider things like the extent to which you have a visual memory of the virtual environment, whether that memory is in colour, the extent to which the memory seems vivid or realistic, its size, location in your imagination, the extent to which it is panoramic in your imagination, and other such structural elements.

6. During the time of your experience, did you often think to yourself that you were actually in the virtual environment?

User Engagement Scale (UES) is a standardised questionnaire used to measure a user's engagement when interacting with an application or system (O'Brien et al., 2018). It consists of 30 items (originally 32), all rated on a 5-point Likert scale, relating to one out of four factors: Focused attention, Perceived usability, Aesthetic appeal, and Reward. When administering the UES the items ought to be randomised. The 30 items in the questionnaire are:

- I lost myself in this experience.

- I was so involved in this experience that I lost track of time.

- I blocked out things around me when I was using Application X.

- When I was using Application X, I lost track of the world around me.

- The time I spent using Application X just slipped away.

- I was absorbed in this experience.

- During this experience I let myself go.

- I felt frustrated while using this Application X.

- I found this Application X confusing to use.

- I felt annoyed while using Application X.
- I felt discouraged while using this Application X.
- Using this Application X was taxing.
- This experience was demanding.
- I felt in control while using this Application X.
- I could not do some of the things I needed to do while using Application X.
- This Application X was attractive.
- This Application X was aesthestically appealing.
- I liked the graphics and images of Application X.
- Application X appealed to be visual senses.
- The screen layout of Application X was visually pleasing.
- Using Application X was worthwhile.
- I consider my experience a success.
- This experience did not work out the way I had planned.
- My experience was rewarding.
- I would recommend Application X to my family and friends.
- I continued to use Application X out of curiosity.
- The content of Application X incited my curiosity.
- I was really drawn into this experience.
- I felt involved in this experience.
- This experience was fun.

User Engagement Scale Short Format (UES-SF) is a shorter version of the UES using only 12 items (O'Brien et al., 2018). Those are:

- I lost myself in this experience.
- The time I spent using Application X just slipped away.
- I was absorbed in this experience.
- I felt frustrated while using this Application X.
- I found this Application X confusing to use.

- Using this Application X was taxing.
- This Application X was attractive.
- This Application X was aesthetically appealing.
- This Application X appealed to my senses.
- Using Application X was worthwhile.
- My experience was rewarding.
- I felt interested in this experience.

4.5 Data Analysis

This section outlines various methods for analysing and evaluating both quantitative and qualitative data.

4.5.1 Bayesian Inference

Bayesian inference is a statistical framework for understanding the world under conditions of uncertainty. It contrasts with similar probability models in that it assigns probabilities to parameters instead of regarding them as fixed (Williamson, 2022, Chapter 1).

The first step of Bayesian inference lies in the concept of *prior probabilities*. These are probability distributions that describe existing knowledge or beliefs about a particular situation before any new data has been considered (Williamson, 2022, Chapter 1). These distributions are usually based on past experience, general understanding, or informed guesses of the situation at hand. When faced with new information, which is manifested in the form of *observations* (new data or evidence), Bayesian inference does not discard the prior knowledge. Instead, it incorporates this new information through a concept called *likelihood*. The likelihood represents the probability of observing the current data given a specific hypothesis (Williamson, 2022, Chapter 1). Thus, it describes how well a particular hypothesis aligns with the observed evidence.

To update prior beliefs using the observations, *Bayes' theorem* is employed (Williamson, 2022, Chapter 1). The theorem combines the prior probability distribution and the likelihood to arrive at a new distribution called the *posterior probability distribution* (Williamson, 2022, Chapter 1). This outcome represents the updated belief about the situation after considering both the prior knowledge and the new observations. In brief, the posterior probability distribution states how likely different possibilities are in light of new evidence and provides a more refined and data-driven understanding using probabilities. Bayesian inference is also iterative which allows for continuous learning and refining understanding as observations arrive, rendering it valuable for handling uncertainty in various fields such as interaction design

(Williamson, 2022, Chapter 1).

While Bayesian approaches offer advantages like incorporating prior knowledge and providing rich information, it has a few limitations. One of them is that Bayesian approaches are general in nature which highlights the importance of formulating the problem clearly and well (Williamson, 2022, Chapter 1). This requires a specific skill set which takes time, effort, and experience for researchers to develop (Williamson, 2022, Chapter 1). Additionally, the inherent subjectivity of Bayesian inference, i.e. the initial prior distribution being based on the beliefs of the researcher, necessitates careful justification and clear documentation for proper interpretation of the results (Williamson, 2022, Chapter 1).

4.5.2 Hypothesis Testing

Hypothesis testing (or statistical hypothesis testing) is a method used to determine if the data is sufficient to support a given hypothesis (Anderson et al., 2024). This is done by defining a null hypothesis and an alternative hypothesis (the opposite of the null hypothesis). If the null hypothesis can be rejected, the alternative hypothesis can be concluded to be true through statistical inference (Anderson et al., 2024).

In order to determine whether the null hypothesis can be rejected, a statistical procedure is employed. There are two categories of these: parametric and non-parametric (MacKenzie, 2013, Chapter 6). Which category to choose is determined by the type of data which is to be analysed. MacKenzie (2013, Chapter 4) presents four scales of measurement: nominal, ordinal, interval, and ratio. Non-parametric tests can be used on data of either scale, whereas parametric tests can only be used on interval- and ratio data (MacKenzie, 2013, Chapter 6).

Other than the form of the data, the form of the experiment itself also affects which specific tests should be used. This is primarily decided based on the number of IVs and how test conditions are assigned (MacKenzie, 2013, Chapter 6). As described in section 4.2, a within-subjects design refers to each participant being tested on each condition of the IV whereas a between-subjects design refers to when the participants are divided into groups and each group has one assigned test condition.

The most common procedure for hypothesis testing is the analysis of variance (ANOVA) which determines whether an IV had a significant impact on a dependent variable (DV) (MacKenzie, 2013, Chapter 6). ANOVA can be adapted to the design of the experiment. For instance, a ”two-way ANOVA” refers to running ANOVA on an experiment with two IVs (MacKenzie, 2013, Chapter 6).

4.5.3 Grounded Theory

Grounded theory’s core concept is for theory to develop through an iterative process of systematic analysis and interpretation of empirical data. The process alternates between data gathering and data analysis, letting one inform the other. The process is repeated until no new insights surface, and the theory becomes thoroughly developed.

When analysing the data, it is crucial to maintain a balance between impartiality and awareness. This will ensure accurate interpretations and help identify relationships in the data (Sharp et al., 2019, Chapter 9). The aim of the analysis is to identify and define the properties and dimensions of what is referred to as *categories* in grounded theory (Sharp et al., 2019, Chapter 9). This is done through three different coding processes, namely open-, axial-, and selective coding. Open coding involves discovering categories and their properties and dimensions in the data. Open coding is followed by axial coding which involves further developing and articulating the categories, breaking them down into subcategories. Finally, there is selective coding which involves structuring the categories to form a cohesive backbone of the theory (Sharp et al., 2019, Chapter 9).

4.5.4 Interaction Analysis

Sharp et al. (2019, Chapter 9) describe interaction analysis as a qualitative data analysis method to study and understand human interactions within their environments. It focuses on both verbal and non-verbal communication with other people and objects and uses video recordings to capture activities and analyse the interactions. The processing of the material is inductive and collaborative which means that teams of researchers observe, discuss, and analyse the recorded video footage together. This contrasts with many traditional research methods where individual analyses are done in isolation from other researchers (Sharp et al., 2019, Chapter 9).

The initial step is to create a content log without pre-defined categories which allows themes to emerge organically from the data. Repeated viewing and discussion of the recorded material is encouraged and considered part of the process (Sharp et al., 2019, Chapter 9). Researchers then generate hypotheses about the intentions and motivations of the recorded participants. A central feature of interaction analysis is the so-called *cannabalization* of videos where researchers extract, group, and reclassify events from the recorded material to identify whether a behaviour or action is a robust theme or a one-off incident (Sharp et al., 2019, Chapter 9).

4.5.5 Thematic Analysis

Thematic analysis is a flexible and widely used method for analysing qualitative data and aims to identify patterns or themes in the recorded data (Sharp et al., 2019, Chapter 9). Themes are understood as patterns or topics that emerge from the data that are assumed relevant or surprising in the context of the objectives of the studies. These can encompass a range of elements such as behaviours, events, or actions, and are important for understanding the nuances of the data in relation to the objectives of the study. In thematic analysis, the process always begins with an initial review of the data which is followed by a more systematic search for themes across the data set. Usually, the analysis seeks to determine whether identified themes are universally applicable across the data or if they are isolated instances. Depending on the objectives of the study, themes can be classified either as primary findings or as starting points for further analysis (Sharp et al., 2019, Chapter 9).

Affinity diagramming is a common tool used in thematic analysis which organises and synthesises large amounts of qualitative data, ideas, and insights into a hierarchical and coherent structure based on perceived commonalities (Sharp et al., 2019, Chapter 9). The process of creating an affinity diagram begins with collecting individual ideas, observations, or data points. This is usually done using a physical medium like sticky notes on a whiteboard, table, or paper, but it can also be done efficiently on shared boards on digital devices (Harboe & Huang, 2015). The notes are then sorted by looking for natural groupings based on similarities or shared themes. The process is inherently iterative and it is encouraged to rearrange the notes as the team discusses their insights and interpretation of specific data points.

Affinity diagrams are particularly valued for their collaborative nature as the process of discussing and grouping the notes encourages team members to share their perspectives and insights which leads to a more comprehensive understanding of the data (Geyer et al., 2011). Thus, affinity diagrams can bridge interdisciplinary gaps and ensure that different viewpoints are considered and integrated into the final thematic structure.

5

Planning

This chapter goes through the methods mentioned in chapter 4, motivating which ones will be used in the thesis. Table 5.1 provides an overview of the chosen methods, explaining why and how they will be used.

Table 5.1: A summary of why the methods were chosen and their usage in the thesis

Method	**Why?**	**Usage**
Storyboarding	Visualise how users may interact with the prototype	Prototyping
High-fidelity prototyping	Produce a prototype of near finished quality	Prototyping
Usability test	To assess the ease of use and user-friendliness of the prototype	User tests
Between-subjects design	Provide the experimental design based on examining military experience as a subject variable	Experimental design
Direct Observation in a Controlled Environment	To capture real-time user interactions, behaviours, and reactions	User tests
Semi-structured Interviews	Collect rich qualitative data from users	User tests
Questionnaires	Collect demographic user data	User tests
RTLX, SUS, WS	Collect standardised quantitative data on different aspects of user experience	User tests
Hypothesis Testing	Easily implemented and straightforward method for quantitative analysis	Quantitative data analysis
Thematic Analysis	Analyse rich and qualitative data to find recurring themes in user behaviour, preferences, and experiences	Qualitative data analysis

5.1 Prototyping

Considering the particularities of VR, the prototyping process for high- and low-fidelity prototyping needs to be tailored to the project at hand. That is why a method like paper prototyping will not be used. Modelling a three-dimensional environment using paper prototypes would likely be incredibly difficult and time-consuming, hence losing its purpose as a low-fidelity prototyping method. Looking at prototyping as a spectrum from low to high fidelity, the prototyping process will start considerably closer to high fidelity than would be the case when developing for example a mobile application. Except for storyboarding, which will be used to map out the scenario to be played out in the virtual environment, all of the prototyping will be carried out in Unreal Engine 5.3.2. This approach allows the prototype to be continuously run and debugged on the target platform, the Meta Quest 3. Unreal Engine 5.3.2 is chosen for its support of the most recent features available at the start of the development process. The Unreal Engine version will not be upgraded during the development process to avoid potential complications and conflicts in the project files. The process will still include a stage reminiscent of low-fidelity prototyping consisting of creating and testing a simplified version, serving the same purpose as using Wizard of Oz.

5.2 Study Format

The study will take the form usability testing examining a subject variable as the object of study, rather than an IV. A subject variable is an inherent pre-existing characteristic not manipulated by the researcher such as age or gender (van Peer et al., 2012, Chapter 6). In this case, the subject variable is military experience. Considering this, since participants by definition cannot be part of more than one group, a between-subjects design will be used. When studying a subject variable, one has to be cautious when drawing conclusions from the result. This is because one cannot conclusively say that a potential difference in result is caused by the subject variable (van Peer et al., 2012, Chapter 6). For instance, if one is studying participants in two different age groups, one cannot know what aspects of participants in the groups cause the difference in results. Is it because of a difference in life experience? Or is it because of a difference in mobility? In other words, there are many aspects other than the subject variable itself that may affect the results.

5.3 Sampling

Due to the limited time and resources, and the fact that the main contribution of the thesis is the development of the simulator, non-probability sampling will be used rather than probability sampling. Recruiting participants will thus done in the most convenient way possible. However, as mentioned in section 4.3, one has to be aware of the higher risk of potential biases compared to probability sampling.

Three methods will be used for recruiting participants: convenience, snowball, and quota sampling. Quota sampling will be used to recruit at least six participants from the two groups, military and non-military. Convenience and snowball sampling will then be used to facilitate the recruitment to reach those quotas.

5.4 Interviews

Unstructured interviews can yield a lot of rich data to be collected but require a certain skill from the interviewer to keep the discussion focused and avoid taking too much time. Focus groups tend to share this disadvantage with unstructured interviews but also have the downside of requiring the test subjects to be present in the same location, which can be tedious and difficult to arrange. Structured interviews are great for staying on topic but do not generate the same quantity of rich data needed for the analysis. Thus, the form of interview best suited for the thesis is a semi-structured interview because it provides rich data while giving the interviewer the tools necessary to stay on topic and explore the specific areas of interest. These properties align particularly well with this project because previous studies on related topics have been conducted as can be seen in section 2.3, providing a frame of reference for relevant areas that could offer valuable insights. In order to allow for transcription and subsequent qualitative data analysis, the interviews will be recorded using audio.

5.5 Observation

Because the prototype requires specific hardware to run, and because user testing generates data that users cannot consciously express or communicate themselves, the most appropriate form of observation is direct observation in a controlled environment. Indirect observation would be unfeasible for the study since the nature of the prototype requires that any user test be conducted in a controlled environment. Direct observation allows researchers to closely observe and gain insight into the behaviour of each participant. This method enhances the subsequent interviews by allowing the observed behaviours to inform and guide the semi-structured interview questions. As a result, a more complete understanding of the users' experiences and reactions can be gained during testing. In order to allow for post-test observations, every user test will be recorded using video and audio.

5.6 Questionnaires

In the context of user tests, questionnaires may not be suitable since the number of participants is small and participants must be present to test the prototype. Therefore, an interview can be conducted on-site after the user test instead of distributing a questionnaire. Conducting interviews also eliminates the risk of participants forgetting to complete the questionnaire. However, collecting demographic data and standardised measurements on different aspects of user experience is still highly relevant. Therefore, one questionnaire to collect demographic data and three standardised questionnaires will still be used to collect quantitative data with regards to the users' experience in the simulator. Namely, RTLX, SUS, and WS.

Based on TAM and CLT, see sections 3.2 and 3.3, as well as the particularities of VR, three main areas of consideration were identified when it comes to aspects of user experience for the simulator:

- Workload
- Usability
- Presence

In order to measure labour intensity, both the NASA-TLX and RTLX are good fits. The RTLX will be used in favour of the NASA-TLX since it is less time-consuming, easier to implement, and fulfils the same purpose as the NASA-TLX with no notable difference in results (Hart, 2006).

The UES or UES-SF both fall somewhere in between usability and presence so neither was deemed a good fit. Instead, the SUS and WS will be used to directly cover those areas. The reason WS was favoured before the SUS-Q is because it is more extensive and well-established. At the time of writing, Witmer and Singer (1998) has over 7967 citations whereas Usoh et al. (2000) has 1162 according to Google Scholar.

5.7 Quantitative Data Analysis

Bayesian inference is often considered a strong alternative to hypothesis testing because of its ability to incorporate and modify prior beliefs. However, formulating a clear problem statement for it can be challenging and the integration of prior beliefs can be subjective. Instead, hypothesis testing will be used as it enables systematic evaluation and validation of specific assumptions. It is an established and comparatively easy method to implement in quantitative data analysis (Ortiz et al., 2010).

5.8 Qualitative Data Analysis

The study's qualitative data includes participant interactions with the prototype and data collected from interviews and questionnaires. While grounded theory is valued for its depth and richness in theory development, its time-consuming nature renders it unsuitable for this project given the thesis's time frame. There are similar problems with interaction analysis. It too is very time-consuming, and it involves interpreting verbal and non-verbal cues which can be difficult and highly inaccurate without a lot of experience. Therefore, thematic analysis will be applied as it meets the needs of small-scale qualitative data analysis while allowing for systematic identification and interpretation of recurring themes through the interviews.

6

System Design

In this chapter, the development tools used to create the VR prototype are detailed. The process and iterations for developing the prototype are also described.

6.1 Development tools

This section introduces all tools that were used during the prototyping process. This includes all of the hardware devices and software that were used at different points to develop the prototype.

6.1.1 Hardware

The VR system for which the prototype was developed is the Meta Quest 3. The Meta Quest 3 is a modern VR system developed by Meta that was released in October 2023 and contains a range of features that rendered it suitable for the prototype (Meta, 2024b). The most important feature is its potential to run software locally on the device, completely detached from external hardware. Another convenient feature of the Meta Quest 3 is its spatial tracking method. Many VR systems such as the HTC Vive Pro 2 require the user to set up external trackers in the physical space to process spatial data from the HMD and controllers (HTC, 2024). The advantage of external trackers is their ability to always have clear spatial data regardless of whether the controllers are in the HMD camera's line of sight. However, the biggest disadvantage is that it adds another component to the system that needs to be installed in the room before using the hardware. Instead, the Meta Quest 3 uses two cameras in combination with other sensors for spatial tracking, rendering the need for external trackers unnecessary (Meta, 2024b). The user must instead set a virtual boundary in the physical room following the instructions of the operating system. This boundary serves as the physical space for each application the user uses and within which they can move freely.

On the front of the HMD, the Meta Quest 3 has two cameras that allow the user to see a real-time view of the physical room around them in the startup interface or in applications that support it. The cameras will also activate should the user cross the boundary they previously set. This feature is called pass-through and creates a mixed reality experience when it is used (Meta, 2024a). The HMD also has a set of speakers for each ear that creates a stereo effect similar to headphones. Should

Figure 6.1: The Meta Quest 3 HMD with the two frontal cameras visible.

the user want to use headphones instead, they can either connect a pair of wired headphones to the headphone jack on the side of the HMD or connect a pair of wireless headphones through Bluetooth.

Hand tracking on the Meta Quest 3 is achieved using two methods. The first method is to use the cameras on the HMD to track the hands and fingers, allowing them to be used for interacting with menus and the virtual environment. The second method is to use the included motion controllers to track the hands and interact with the virtual environment as shown in figure 6.2. There is one controller designed for each hand and each controller has a few buttons with different purposes. The functionality of every button varies from application to application and the following descriptions of every button's functionality will thus be general usage notes.

On the side of the controller is the grab trigger which the user can use to grab objects in virtual environments. On the front of the controller is a trigger for the index finger that is used to interact with menus in the graphical user interface (GUI). The trigger is also used to interact with some objects in virtual environments that do not require grabbing. On the top of each controller is a joystick and three thumb buttons. The joystick can be used for anything that involves direction such as looking around or moving in virtual environments when the physical space is inadequate in size. The three thumb buttons can be split into two categories: Two convex and one concave. The two convex buttons are chiefly used for interacting with menus and prompts in virtual environments whilst the one concave button is used to open functional menus or activate debug functionality such as resetting the position of menus.

Figure 6.2: The two hand controllers for the Meta Quest 3. The index finger triggers are not visible from this angle.

The HMD contains all of the hardware required to run software locally. It has two displays with a resolution of 2064x2208 pixels and a frame rate of up to 120 Hz each. The operative system that the HMD runs is a version of Android adapted for virtual reality. However, the HMD itself contains limited interaction functionality. It has a power button that is used to power the HMD on and off. It also has a volume button used to control the sound volume of the speakers on the HMD. The volume button can also be used for limited interaction with menus in case the hand controllers are out of operation or out of reach. There is also an adjustment wheel underneath the HMD which is used to adjust the lens spacing should the visual clarity be insufficient. There are two straps on the HMD that are used to fit the HMD as well as possible to the user's head. One of the straps goes over the scalp whilst the other goes around the skull horizontally. Once both straps are tightened, the HMD should fit snugly on the head without the risk of falling off due to quick movements. All in all, the HMD weighs about half a kilogram (Meta, 2024b).

6.1.2 Software

The VR prototype was developed using a combination of various tools. The development computers ran Windows 10, and the game engine and integrated development environment used for the prototype was Unreal Engine 5.3.2. Unreal Engine was chosen for a few reasons.

Firstly, Unreal Engine provides stable support for development on the Meta Quest 3 which ensured that the game engine and VR system remained compatible throughout the development process. This compatibility also offered a reliable foundation and allowed the focus to be directed toward creating and refining the VR experience without concerns about integration issues or potential disruptions from software updates. Secondly, Unreal Engine is the game engine that Tactisim use for their products which enables integration of their assets when necessary and facilitates effective communication with their developers. Thirdly, Unreal Engine provides a preview function in the editor that allows the prototype to be viewed using rendering engines optimised for different platforms. This feature was important as most graphical rendering in Windows 10 uses a different graphics API to Meta Quest 3, which could make it difficult to predict how models would behave once the prototype was transferred and run on Meta Quest 3.

Initially, GitHub was chosen as the version control system. However, it was soon abandoned because it only offered 1GB of online storage space which proved insufficient for the project's needs. The project was subsequently migrated to GitLab, which provided 10GB of online storage space in its free version. This allowed for more extensive version control and management of larger files without the constraint of limited storage. To ensure that no single developer could overwrite the main branch without approval, the repository's main branch was protected from direct manipulation. An internal system was implemented where a functional version of the prototype was maintained in the main branch. All development occurred in branches derived from the main branch, which were then merged back into the main branch once a function or feature was considered complete and had passed a series of tests in the VR system.

To expedite the development process, the decision was made early on to avoid creating models from scratch. Instead, existing models would be sourced from various repositories and selected based on their compatibility with the game engine and how well they met the end goal of the application. In addition, it was decided that assets would generally be excluded should they exceed 5MB. This decision was made because of three reasons. Firstly, large assets would eat up space quickly in the GitLab repository which constrains the number of total assets in the prototype. Secondly, larger assets tend to hurt performance in Unreal Engine since they have more polygons to render. Lastly, larger assets render commits and pushes into the GitLab repository longer to complete, slowing down the development process even more. Nonetheless, a few exceptions were made to the 5MB rule, notably when no other suitable asset was found.

One of the websites used was Sketchfab (2024) which is an asset website that provides 3D models for various purposes including VR which naturally became the focus point of this project. To ensure that there would not be any copyright issues with the assets, the prototype only used models which were contained under a Commons (2024) license on SketchFab. Another source of assets used during the prototype development was Quixel Bridge, an asset library developed by Quixel (2024) built into Unreal Engine. The biggest advantage of Quixel Bridge is that all of its assets are free to use as long as they are not used outside of Unreal Engine. Quixel Bridge was used for environmental assets such as the textures for the grass and dirt roads but also to add details such as rocks and logs. Another crucial feature of Quixel Bridge was that one could choose in which resolution to download the texture before importing it into the Unreal Engine editor. The concern regarding whether a texture was too big was thus rendered unimportant. A disadvantage to Quixel Bridge is that it solely contains assets based on megascans, which are models based on high-resolution photographs, limiting its use in any other area than environmental assets.

6.2 Development process

This section dives deeper into the development process in Unreal Engine.

6.2.1 Scenario development

The scenario integrated into the prototype was provided by Tactisim and included a squadmate walking along a road in an open space and suddenly stepping on a landmine. The explosion is powerful enough to amputate the soldier's foot and part of their leg up to and including the ankle, immediately starting a catastrophic bleed. The user must now try to save their squadmates from bleeding out. Tactisim's scenario was constructed as a story with pauses for multiple-choice questions, making the possible actions clear to the user. The correct sequence of actions according to the scenario description is then to:

1. Assess the situation
2. Slowly walk to the squadmate while cautiously avoiding other mines
3. Reach the squadmate
4. Apply a tourniquet to the lower or middle left leg
5. Secure the twist pin on the tourniquet
6. Fold over the tourniquet strap
7. Get a pen to sign the tourniquet
8. Write the date on the tourniquet in the format DDHHmm (Day, hour, minute)
9. Call for pickup of the wounded squadmate and report the event

The scenario description included both incorrect and non-optimal choices for various actions which contributed to the storytelling becoming less linear. An incorrect choice is a significant error that immediately terminates the scenario. For instance, choosing to hurry to one's squadmate instead of first assessing the situation and carefully approaching them results in the user stepping on a mine and ending the scenario instantly. In contrast, a non-optimal choice represents a minor error that, while not immediately fatal to the scenario, could be executed more efficiently. For example, when writing the date on a tourniquet, the scenario includes an option for an incorrect date format. Selecting this non-optimal choice does not end the scenario instantly, but it indicates a non-optimal action that could have been performed better by writing the date in another format.

This scenario description was provided as a YAML file, a file format intended to make data serialisation straightforward and human-readable. The YAML file was converted to a flow chart (see Appendix A) which was used during the development of the prototype since it was more efficient to find the part of the scenario being worked on.

6.2.2 Prototyping

The development of the prototype took around two months from start to finish. Initially, time was spent researching the functionality of Unreal Engine and getting accustomed to working with the engine in combination with version control. After that, the level was constructed using one of Tactisim's levels as inspiration. The existing level could not be used directly because its models were too high in resolution which slowed down the performance of Unreal Engine considerably and even rendered it completely unresponsive at times. Tactisim's level was also adapted for a desktop environment which utilised graphical rendering features that the Meta Quest 3 does not support, rendering the conversion between the desktop and VR graphical APIs tedious and time-consuming.

After constructing the level, the implementation of the VR scenario commenced. To maintain focus on the most critical aspects of the scenario, the initial integration prioritised the path that included only the correct options from the scenario description while adapting the multiple-choice actions to suit the VR platform. The implementation began by spawning the user into the level. Subsequently, the squadmate and its model were developed and featured a predefined sequence in which the squadmate walked onto a landmine, causing it to detonate and amputate the squadmate's leg. Beyond this point, no additional prerecorded sequences were included; the user was thus required to progress naturally through the scenario without any guidance from the interface.

It was determined that multiple-choice actions were not optimal for a VR experience. Consequently, efforts were made to eliminate as many of these multiple-choice actions as possible. For example, when assessing the situation after a squadmate has detonated a mine, the original description required the user to look around for a bit before proceeding. It was theorised that users would instinctively survey their surroundings upon entering the level, and it was therefore decided not to implement any mechanisms to enforce this action, such as disabling virtual locomotion for the first few seconds of the simulation. This solution relied on the intuitive behaviour of users to create a more seamless and immersive experience and avoided unnecessary interruptions that could detract from the realism and flow of the scenario.

Some multiple-choice actions had to be simplified due to the limitations and characteristics of the platform, the Meta Quest 3. With the exception of vibration, there is a complete lack of tactile feedback, effectively leaving out the sense of touch completely. Furthermore, the use of hand controllers inhibits the ability to exercise fine motor skills that require finger precision. Consequently, the process of applying the tourniquet was simplified to placing the tourniquet on the leg and pressing and holding a single button to tighten it, replacing the previous actions of steps 4, 5 and 6 as seen in the list in section 6.2.1. Similarly, the signing of the tourniquet was simplified to pressing and holding a single button, replacing the previous actions of steps 7 and 8 as seen in the list in section 6.2.1. One additional contributing factor to this limited implementation was the overwhelming time constraints of the development process for the prototype.

To allow for movement unrestricted of physical space, some form of virtual locomotion was deemed necessary. Coomer et al. (2018) and Bond and Nyblom (2019) both evaluate four different virtual locomotion techniques. Whilst Coomer et al. (2018) identify *arm-cycling* as having advantages to other techniques, Bond and Nyblom (2019) identify *teleportation* as the potentially most fitting technique for interactive environments specifically. Ultimately, teleportation was chosen for the prototype with the motivation that real-world movement would be used for moving short distances and teleportation for long distances. In other words, teleportation to get to a general area followed by real-world movement to move around in that area.

One challenging aspect of the prototype design was directing the user through the scenario. Some parts of the scenario were deemed too unclear without enough guidance and as a result, graphical hints were implemented in some spots to help the user find objects essential for progressing through the scenario. An example of this is spawning the tourniquet into the world. To grab the user's attention, the tourniquet is only spawned once the user has entered an area encapsulating the soldier. Along with spawning the tourniquet, a text above the tourniquet is also spawned, alerting the user further that the tourniquet is there. Directing the user this way lowers extraneous and germane cognitive load, concepts outlined in section 3.3, clarifying the next step of the scenario at the expense of realism.

After the complete implementation of the correct path through the scenario, attention turned to incorporating the incorrect options into the path. This process generally mirrored the approach used for the correct path as the actions for both correct and incorrect options were typically similar. For example, when applying the tourniquet to a part of the leg, the user only needs to position the tourniquet at a specific spot and release the grab button. This action automatically mounts the tourniquet to one of two predefined locations on the leg either near the amputation or higher up on the thigh. The technical implementation of these options was essentially the same which streamlined the process of implementing the incorrect options.

The final addition to the scenario is the inducement of stress. Binsch et al. (2021) and Kleygrewe et al. (2023), emphasise the importance of inducing stress when learning casualty care. To create a sense of urgency in the scenario, if the user has not applied the tourniquet within 60 seconds of the squadmate triggering a landmine, the squadmate bleeds out. No indication of how much time is left, other than the size of the pool of blood is presented to the user, with the motivation of it being unknown is an additional source of stress.

When the implementation of the correct and incorrect paths was completed, the focus shifted toward testing the prototype to identify and document as many apparent bugs as possible. This initial phase of testing was crucial for ensuring at least some reliability of the VR scenario before progressing to pilot testing. During these tests, any bug that was found was noted and analysed to determine its impact on the user experience. To manage this process efficiently, a list of prioritised bug fixes was created. The list ranked the bugs by their severity with the highest priority issues at the top and lower priority ones further down. The process of bug-fixing thus entailed working through this list from top to bottom which ensured that the most urgent issues were addressed first with lower-priority bugs being taken care of later. Throughout this process, it was common for the solution of one bug to inadvertently create another bug. When this happened, the newly introduced bug was immediately prioritised and fixed before continuing down the priority list. This ensured that any cascading problems were promptly managed and prevented bugs from compounding which could affect the pilot tests negatively.

6.2.3 Pilot testing

Once the list of bug fixes was finished, the next step was to conduct a series of pilot tests. The pilot tests were chiefly carried out to help identify issues with the prototype or the testing process itself. These tests spanned over a week during which the early version of the application was tested on five participants. These participants were recruited using convenience sampling since it allowed the pilot tests to be performed on short notice.

The participants were asked to go through the simulator multiple times. If one participant got stuck at a specific point in the simulator, they were given either a hint to guide them in the correct direction to resolve the situation or a direct instruction if their focus was directed at something other than the scenario itself.

Following each pilot test, participants were asked questions about their experience. This provided the first opportunity to gather outsider perspective feedback on the functionality of the prototype. Additionally, the short conversations that followed aided in shaping the questions for the interviews in the final user tests.

The pilot tests were held at various locations to assess the suitability of each location for the final user tests. There were a few features that were looked for at every location to find the optimal place for user testing. The main feature that was looked after was space since the participants needed a lot of room to be able to move somewhat freely within the simulation. Space was also essential to give the participants a spacious boundary which warned them whenever they were about one meter away from a wall or obstacle in the room. This elevated the space requirement considerably but prioritising it was deemed necessary to ensure that no participant would collide with a wall or other obstacles that were present in the room such as tables, chairs, projectors and so forth. Another aspect that was considered when choosing a location was privacy. To make the participants comfortable, rooms with little to no possibility of passersby looking in were preferred. The last aspect considered was booking accessibility as these rooms would at times have to be booked with little notice when finding research participants.

6.2.4 Second iteration

After completing the pilot tests, the feedback and data collected were analysed to identify necessary fixes and changes before the user tests began. This primarily involved resolving bugs with features that did not behave as expected and modifying parts of the user interface for better clarity. One of the most significant changes was introducing a starting level separate from the main scenario. In this starting level, users can take the time they need to familiarise themselves with the controls that they will need to finish the scenario. This level is separate from the level in which the scenario takes place, and the levels do not overlap at all. The second iteration was the last development step in the thesis project and the resulting prototype is therefore the final prototype described in chapter 7.

7

Final Prototype

This chapter describes the final prototype that resulted from the prototyping process including its functionality, assets, and more.

7.1 Simulator

The final prototype deemed fit for user testing was completed after the wrap-up of the second iteration. The final prototype consists of two levels with functionality isolated from each other. The starting level features a large, flat grassy area that allows users to move freely by walking in the physical space outside the simulator as shown in figure 7.1. Another way of moving around in the level is teleportation. Teleportation is performed by pressing the joystick on the right controller forward, displaying a visual indicator that shows where the user will teleport to after releasing the joystick back to its original state as shown in figure 7.5. Upon releasing the joystick, the user is immediately teleported to the indicated location.

The starting level also contains objects such as rocks, cubes, and rifles for users to practice grabbing, dropping, and throwing objects in the virtual environment. After one minute in the starting level, a menu for starting the scenario appears, prompting the user to begin by pointing and clicking the start button as shown in figure 7.2. Clicking the button removes the starting level and spawns the user into the scenario level, starting the scenario instantly.

The scenario starts with the user spawning onto a dirt road in a forest. The environment includes a few objects that the user can see, such as a white van, a log of a fallen tree blocking the road ahead, some small rocks on their right side, and a few trees forming a forest on the sides of the dirt road as shown in figure 7.3. The user can instantly start to move around to inspect and interact with the environment, but attention is immediately attracted to another squadmate walking towards them. After approximately six seconds, the squadmate commits a critical mistake and steps on a landmine that blows up, knocking the squadmate backwards and amputating his leg. The amputated limb now starts bleeding catastrophically and the user has precisely one minute to tighten the tourniquet around the leg.

As a first step, the user must now cross the minefield. The landmines on the field are spread out in a random pattern, but the tip of every landmine sticks up through the

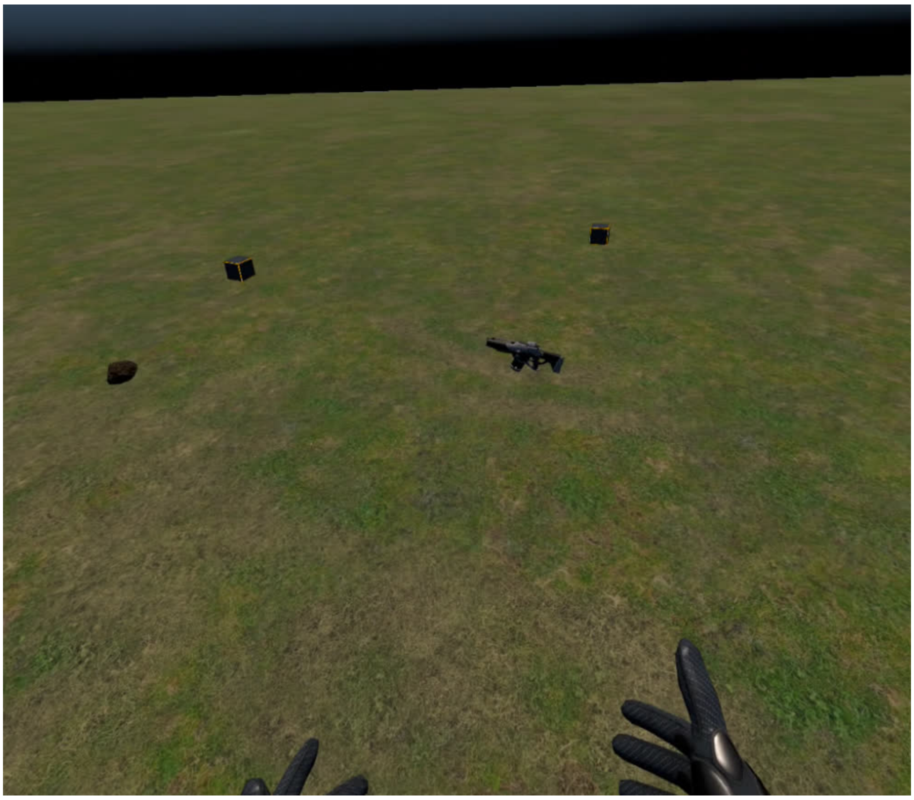


Figure 7.1: The starting level of the simulator where the user can get familiarised with the controls.

Figure 7.2: The menu used to start the scenario.

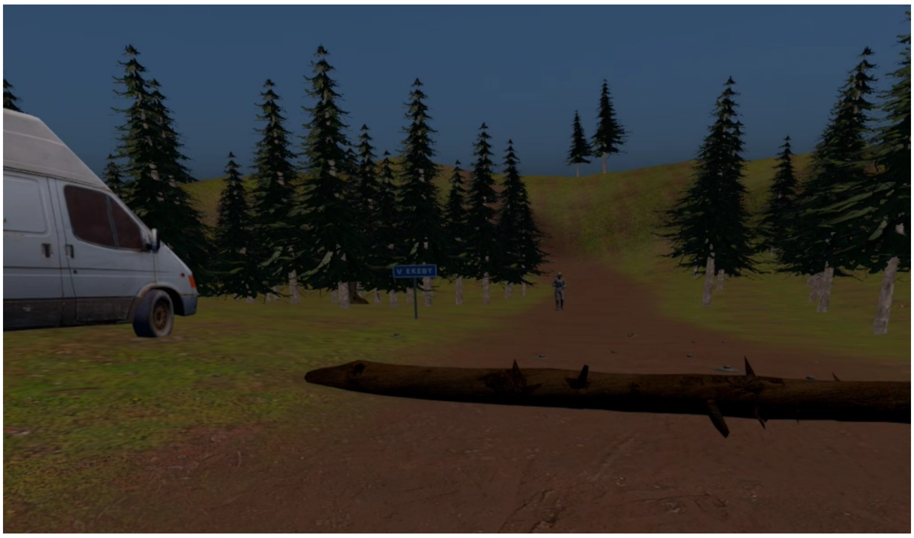


Figure 7.3: The level where the scenario takes place.

dirt road, giving the user a visual clue of where not to step. Due to constraints with physical space, the user must now find a path through the minefield by teleporting. If the user teleports in a way that a line between the previous and new location intersects a mine, the mine explodes and the scenario ends instantly. Ending the scenario brings up a restart menu that prompts the user with the committed error and displays an interactive button that restarts the scenario from the beginning as shown in figure 7.4. The menu is always spawned facing the user, making it easy to spot and interact with.

Figure 7.4: The menu after the user steps on a landmine.

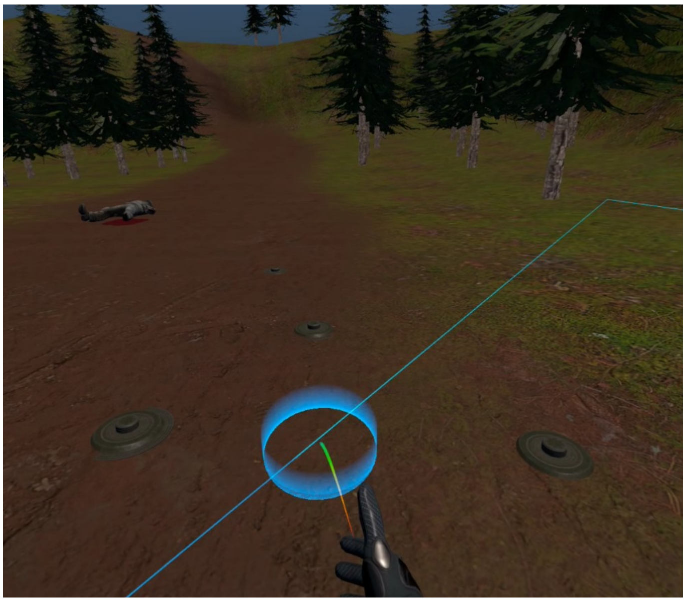


Figure 7.5: User traversing the minefield using teleportation.

When the user has traversed the minefield without triggering a landmine, they now find themself near the wounded squadmate. When they enter an invisible trigger box surrounding the squadmate, a tourniquet spawns beside the squadmate with an accompanying text that reads “tourniquet”. The user can now interact with the tourniquet by grabbing it and picking it up off the ground. The grabbing action is performed by squeezing either the right or left controller whilst intersecting the tourniquet, sticking it to the squeezing hand and letting the user move it as they please. The tourniquet can now be mounted to two spots: either close to the amputation or higher up on the thigh. Both options let the user progress through the scenario, but placing it close to the amputation is the only correct option. The user mounts the tourniquet to the leg by holding it where they want to put it and releasing it, see figure 7.6. This automatically mounts the tourniquet to the leg without the user having to unfold it and thread the strap of the tourniquet around the leg. Should the user not be content with the placement, they can grab it again which dismounts the tourniquet from the leg and the user can now place it in a new location.

Once the tourniquet is mounted, the user is prompted to tighten it. This is done by holding the A thumb button on the right controller, starting a progress bar that takes five seconds to complete, see figure 7.7. Should the user release the A thumb button before the progress bar is filled, the action is reset and the user must restart the tightening process. While tightening the tourniquet, a sound can be heard that indicates to the user the tourniquet is being tightened. After the progress bar is complete, the tourniquet is tightened and can no longer be unmounted and moved to another spot. At the same moment, the user is prompted to sign the tourniquet. The tourniquet can be signed using two formats: one short and one long. The options are displayed as texts hovering above each hand that always face the user. The user must now follow the same holding procedure as when tightening the tourniquet,

Figure 7.6: User about to mount the tourniquet just above the amputation.

taking five seconds to complete. However, since there are two options, the user can hold the X thumb button on the left controller or the A thumb button on the right controller. While signing, a scribbling sound can be heard, indicating to the user that the tourniquet is being signed. After the tourniquet is signed, a handwritten note can be seen in either format that shows the exact time the tourniquet was applied as shown in figure 7.8.

The last step of the scenario is to report the emergency. After the tourniquet is applied, a walkie-talkie is spawned beside the squadmate as shown in figure 7.8. To report the emergency, the user must grab the walkie-talkie which plays a short voice line and instantly ends the scenario. After ending the scenario, an interface that includes information about how well the user performed the task is displayed as seen in figure 7.9. Three metrics are shown: task completion time (TCT), whether the user applied the tourniquet in the correct spot, and whether the user signed the tourniquet using the correct date format. The user is prompted to restart the scenario using the same procedure as when stepping on a landmine described earlier in this section.

The scenario can be restarted several times after entering, meaning the user can take the time needed to get habituated to the simulator. To close the simulator completely, the user can either close it through the operating system or by opening a developer menu and pressing the exit button. Should the user want to exit the simulator immediately, they may also remove the VR HMD while it is running.

Figure 7.7: User in the process of signing the tourniquet.

Figure 7.8: The walkie-talkie immediately after spawning next to the soldier.

Figure 7.9: The end screen after reporting the emergency. In this scenario, the tourniquet was applied correctly but was signed incorrectly.

7.2 Assets

Most of the environmental assets used in the simulator emanated from Quixel Bridge since it was the most convenient option most of the time, being built into Unreal Engine. The assets were usually realistic in detail which rendered them suitable for the environment. SketchFab was used for most objects that did not pertain to the environment and thus could not be found in Quixel Bridge. Examples of this include the tourniquet and the squadmate.

8

Study

The study was designed and conducted as a usability test, observing the participants interact with the prototype, using a between-subjects design with military experience in place of an IV. The purpose of using military experience in place of an IV is two-fold. Firstly, since there was no time to perform a follow-up study and measure learning retention, comparing those with previous knowledge in the field and those without might reveal something about how effective the simulator is at teaching the material. Secondly, considering the diverse use cases for the simulator - ranging from educating novices to enhancing specialised skills for those with prior knowledge - this comparison could uncover potential shortcomings and flaws in the simulator's performance for either group.

Each test included 10 trials, meaning each participant went through the scenario in the simulator ten consecutive times. For each test, the following three points were recorded:

- Number of unsuccessful trials before the first successful trial, or "failures before success" (FBS).
- Number of successful trials.
- Total time spent in the simulator.

To determine the participant's performance in each trial, the TCT and eight stimuli were considered. The stimuli were added up to provide a single score from zero to eight. The stimuli were defined as follows:

1. Did not throw rocks.
2. Did not step on a landmine.
3. Picked up the tourniquet.
4. Applied the tourniquet.
5. Applied the tourniquet correctly.
6. Signed the tourniquet.

7. Signed the tourniquet correctly.
8. Picked up the walkie-talkie.

Furthermore, three different questionnaires: RTLX, SUS, and WS as well as semi-structured interviews were used as follow-ups to the trials to gain insight into the participants' own experience of the simulator. The WS questionnaire was altered to include only 30 of its 32 items. Question 17 was excluded based on being irrelevant since it asks about touch which was not integrated in the simulator. Question 24, *How distracting was the control mechanism?*, was excluded based on possible ambiguity since *the control mechanism* could refer to the HMD, the hand controllers, the VR system as a whole, or how the movement was implemented in the simulator.

The first part of the semi-structured interviews consisted of five predetermined open questions related to their experience using the system. The questions were asked in the following order:

1. How would you describe your experience in the scenario?
2. Were there any details in the scenario that stood out? In what way?
3. Describe step by step, the correct sequence of events for the scenario.
4. What is your view on using this type of tool for training and educational purposes?
5. How was your experience answering the questionnaires?

The second part of the interview focused on letting the participants elaborate and clarify their answers to some of the questions in the questionnaires.

8.1 Participants

The participants (n = 14) in the study (2 female, 12 male) were recruited using a combination of convenience, snowball, and quota sampling. Six of the participants had military experience. The data gathered from one (P3) of the eight participants without military experience was excluded from the study because the VR system malfunctioned during their test. This results in n = 13 participants, 2 female and 11 male, and 6 with military experience and 7 without.

8.2 Hypotheses

In accordance with hypothesis testing, 28 null and their respective alternative hypotheses were formulated in order to perform the quantitative analysis of the study, determining whether military experience (subject variable replacing IV) has any effect on experience or performance in the simulator (DVs):

Test Hypotheses:

H_{0_1} : Military experience has no effect on RTLX score
H_{1_1} : Military experience has an effect on RTLX score
H_{0_2} : Military experience has no effect on SUS score
H_{1_2} : Military experience has an effect on SUS score
H_{0_3} : Military experience has no effect on WS score
H_{1_3} : Military experience has an effect on WS score
H_{0_4} : Military experience has no effect on FBS
H_{1_4} : Military experience has an effect on FBS
H_{0_5} : Military experience has no effect on the number of successful trials
H_{1_5} : Military experience has an effect on the number of successful trials
H_{0_6} : Military experience has no effect on the total time spent in the simulator
H_{1_6} : Military experience has an effect on the total time spent in the simulator
H_{0_7} : Military experience has no effect on the average score
H_{1_7} : Military experience has an effect on the average score
H_{0_8} : Military experience has no effect on the average TCT
H_{1_8} : Military experience has an effect on the average TCT

Trial Hypotheses:

For $i = 1, \ldots, 10$ (each trial):

$H_{0_{8+i}}$: Military experience has no effect on the score from trial i
$H_{1_{8+i}}$: Military experience has an effect on the score from trial i
$H_{0_{18+i}}$: Military experience has no effect on the TCT from trial i
$H_{1_{18+i}}$: Military experience has an effect on the TCT from trial i

8.3 Procedure

A user test began with the participant reading through an informed consent form, see Appendix B. The informed consent form contained information about what the participant was going to experience, how the different parts of the test worked, and clarified the participant's right to withdraw from the study at any point and for any reason. Once the participant accepted the informed consent by putting a cross or their name in the "I consent" field, the study began with a quick explanation of the controls and mechanisms of the Meta Quest 3. The participant was then asked to mount the HMD, receiving help if necessary. Afterwards, the participant was given the controls to begin interaction with the VR software.

The first thing the participants saw when entering the simulator was the starting level in which the participants had time to familiarise themselves with the controls. When the participant proclaimed that they felt ready or seemed adequately proficient with the controls and surroundings, they were given a short explanation that they were going to go through the scenario ten times regardless of the outcome and that there would be little to no communication with the researchers until they finished all ten trials. After they had understood, they were allowed to enter the scenario and start the simulation. After finishing the simulation, the participant was asked to remove the HMD and was then allowed to adapt a bit to having the HMD removed. They were then asked to sit down to start filling out the questionnaires. After completing the questionnaires, the participant went through a semi-structured interview as described earlier in the chapter.

9

Result

The items in questionnaires RTLX and SUS were scored on a 5-point Likert scale and the items in WS were scored on a 7-point Likert scale. Positively formulated questions were scored by subtracting one from the participant's answer whereas negatively formulated questions were scored by subtracting the participant's score from the maximum, i.e. 5 or 7 depending on the scale. Each questionnaire was scored by averaging the score from their items and then mapping it to a score between 0 and 100.

Table 9.1 presents demographic data about the participants, their scores from the questionnaires RTLX, SUS and WS, and their highest achieved score from all trials. The overall means and standard deviations for the questionnaires are presented in table 9.2 and the means and standard deviations specific to each of the groups (military and non-military) are presented in table 9.3. The mean scores and standard deviations specific to each of the groups (military and non-military) for each trial are visualised in figure 9.1.

According to Soegaard (2023), a SUS score above 70 is considered good and a score above 85 excellent. This means that the mean SUS scores (all participants, military group, and non-military group) fall just short of excellent.

According to Hertzum (2021) the mean TLX scores from the study (all participants, military group, and non-military group) fall somewhere below the 10th percentile overall, and below the means for the domains of emergency response and engineering design (46±11 and 38±12).

As for WS, there was neither a standardised method nor convention to score or evaluate the score. Therefore, the WS was scored in the same manner as the SUS and TLX. The main use of the WS was thus to help guide the interviews during the study and highlight specific points mentioned by the participants.

Table 9.1: The demographic data and scores for each participant.

Participant	Gender	Age range	Military	RTLX	SUS	WS	High Score
1	Male	18-24	Yes	4.17	70.00	59.68	8
2	Male	25-34	No	0.00	82.50	65.05	8
4	Male	18-24	Yes	20.83	85.00	68.28	8
5	Male	25-34	Yes	20.83	90.00	70.97	8
6	Female	18-24	Yes	25.00	72.50	59.68	8
7	Male	25-34	No	20.83	72.50	58.60	8
8	Male	18-24	No	25.00	90.00	82.26	8
9	Male	18-24	No	20.83	92.50	85.48	8
10	Male	18-24	No	45.83	85.00	69.89	8
11	Male	25-34	No	20.83	82.50	69.35	8
12	Male	18-24	Yes	33.33	80.00	56.99	8
13	Male	18-24	Yes	20.83	82.50	72.04	8
14	Female	25-34	No	16.66	90.00	75.81	8

Table 9.2: The means and standard deviations of the results from the questionnaires.

Questionnaire	M	SD
RTLX	21.15	11.34
SUS	82.69	7.32
WS	68.78	8.90

Figure 9.1: Mean scores and standard deviation for each individual trial.

9.1 Quantitative data analysis

In accordance with the method of Hypothesis testing, 28 statistical significance tests, each corresponding to one of the hypotheses, were performed using military experience in place of an IV. One instance of Welch's unequal variances t-test with a significance level of 5% ($p < 0.05$) for each of the DVs. Presented below are the results from the tests performed. They are also summarised in Table 9.3.

Table 9.3: The results from the Welch's unequal variances t-tests. Note that there is no significant difference between military and non-military for any of the DVs.

				Military		Non-military	
DV	t	df	p <	M	SD	M	SD
RTLX	0.09	10.67	1	20.83	9.50	21.43	13.49
SUS	1.24	10.20	1	80.00	7.58	85.00	6.77
WS	1.73	10.61	1	64.61	6.57	72.35	9.49
FBS	-0.82	9.90	1	5.00	2.68	3.86	2.27
Successes	0.32	11.00	1	4.33	1.97	4.71	2.29
Tot. Time	1.19	10.66	1	415.83	87.09	485.57	123.94
Avg. Score	0.64	5.68	1	5.68	0.86	5.91	0.24
Avg. TCT	1.04	10.80	1	32.47	6.81	36.46	7.03
T1 Score	-1.55	6.76	1	2.50	1.97	1.14	0.90
T2 Score	-0.15	10.45	1	3.83	3.13	3.57	2.94
T3 Score	0.08	8.01	1	6.33	2.66	6.43	1.62
T4 Score	1.27	9.23	1	4.33	3.33	6.43	2.51
T5 Score	2.34	5.44	1	4.17	3.49	7.57	0.79
T6 Score	-1.41	6.33	1	7.67	0.52	5.86	3.34
T7 Score	1.53	5.17	1	5.50	3.51	7.71	0.49
T8 Score	-0.47	11.00	1	6.50	2.74	5.71	3.25
T9 Score	-1.00	6.00	1	8.00	0.00	7.86	0.38
T10 Score	-1.16	6.00	1	8.00	0.00	6.86	2.61
T1 TCT	0.71	10.38	1	31.50	25.52	41.29	23.63
T2 TCT	1.93	5.84	1	35.00	30.68	60.14	9.62
T3 TCT	1.49	10.25	1	49.83	8.40	58.86	13.25
T4 TCT	0.77	9.02	1	30.67	21.48	38.86	15.64
T5 TCT	0.95	8.79	1	28.00	26.85	40.43	18.79
T6 TCT	-0.95	8.81	1	35.33	10.27	26.57	21.78
T7 TCT	0.41	5.31	1	25.83	20.53	29.29	3.90
T8 TCT	-1.15	9.94	1	28.17	14.37	19.57	12.22
T9 TCT	-1.41	5.86	1	33.17	10.53	26.86	3.34
T10 TCT	-1.24	8.33	1	27.17	3.66	22.61	8.64

There was no significant effect on the RTLX scores, $t(10.67) = 0.09$, $p = 0.93$, despite non-mil. ($M = 21.43$, $SD = 13.49$) attaining higher scores than mil. (M = 20.83, SD = 9.50).

There was no significant effect on the SUS scores, $t(10.20) = 1.24$, $p = 0.24$, despite non-mil. ($M = 85.00$, $SD = 6.77$) attaining higher scores than mil. ($M = 80.00$, $SD = 7.58$).

There was no significant effect on the WS scores, $t(10.61) = 1.73$, $p = 0.11$, despite non-mil. ($M = 72.35$, $SD = 9.49$) attaining higher scores than mil. ($M = 64.61$, $SD = 6.57$).

There was no significant effect on the FBS, $t(9.90) = -0.82$, $p = 0.43$, despite non-mil. ($M = 3.86$, $SD = 2.27$) attaining lower scores than mil. ($M = 5.00$, $SD = 2.68$).

There was no significant effect on the number of successful completions, $t(11) = 0.32$, $p = 0.75$, despite non-mil. ($M = 4.71$, $SD = 2.29$) attaining higher scores than mil. ($M = 4.33$, $SD = 1.97$).

There was no significant effect on the total time spent in the simulator, $t(10.66) = 1.19$, $p = 0.26$, despite non-mil. ($M = 485.57$, $SD = 123.94$) attaining higher scores than mil. ($M = 415.00$, $SD = 87.09$).

There was no significant effect on average score, $t(5.68) = 0.64$, $p = 0.55$, despite non-mil. ($M = 5.91$, $SD = 0.24$) attaining higher scores than mil. ($M = 5.68$, $SD = 0.86$).

There was no significant effect on average TCT, $t(10.80) = 1.04$, $p = 0.32$, despite non-mil. ($M = 36.46$, $SD = 7.03$) attaining higher scores than mil. ($M = 32.47$, $SD = 6.81$).

There was no significant effect on the score from trial 1, $t(6.76) = -1.55$, $p = 0.17$, despite non-mil. ($M = 1.14$, $SD = 0.90$) attaining lower scores than mil. ($M = 2.50$, $SD = 1.97$).

There was no significant effect on the score from trial 2, $t(10.45) = -0.15$, $p = 0.88$, despite non-mil. ($M = 3.57$, $SD = 2.94$) attaining lower scores than mil. ($M = 3.83$, $SD = 3.13$).

There was no significant effect on the score from trial 3, $t(8.01) = 0.08$, $p = 0.94$, despite non-mil. ($M = 6.43$, $SD = 1.62$) attaining higher scores than mil. ($M = 6.33$, $SD = 2.66$).

There was no significant effect on the score from trial 4, $t(9.23) = 1.27$, $p = 0.24$, despite non-mil. ($M = 6.43$, $SD = 2.51$) attaining higher scores than mil. ($M = 4.33$, $SD = 3.33$).

There was no significant effect on the score from trial 5, $t(5.44) = 2.34$, $p = 0.06$, despite non-mil. ($M = 7.57$, $SD = 0.79$) attaining higher scores than mil. ($M = 4.17$, $SD = 3.49$).

There was no significant effect on the score from trial 6, $t(6.33) = -1.41$, $p = 0.20$, despite non-mil. ($M = 5.86$, $SD = 3.34$) attaining lower scores than mil. ($M =$

7.67, SD = 0.52).

There was no significant effect on the score from trial 7, $t(5.17) = 1.53$, $p = 0.18$, despite non-mil. (M = 7.71, SD = 0.49) attaining higher scores than mil. (M = 5.50, SD = 3.51).

There was no significant effect on the score from trial 8, $t(11) = -0.47$, $p = 0.65$, despite non-mil. (M = 5.71, SD = 3.25) attaining lower scores than mil. (M = 6.50, SD = 2.74).

There was no significant effect on the score from trial 9, $t(6.00) = -1$, $p = 0.36$, despite non-mil. (M = 7.86, SD = 0.38) attaining lower scores than mil. (M = 8.00, SD = 0.00).

There was no significant effect on the score from trial 10, $t(6.00) = -1.16$, $p = 0.29$, despite non-mil. (M = 6.86, SD = 2.61) attaining lower scores than mil. (M = 8.00, SD = 0.00).

There was no significant effect on the TCT from trial 1, $t(10.38) = 0.71$, $p = 0.49$, despite non-mil. (M = 41.29, SD = 23.63) attaining higher scores than mil. (M = 31.50, SD = 25.52).

There was no significant effect on the TCT from trial 2, $t(5.84) = 1.93$, $p = 0.10$, despite non-mil. (M = 60.14, SD = 9.62) attaining higher scores than mil. (M = 35.00, SD = 30.68).

There was no significant effect on the TCT from trial 3, $t(10.25) = 1.49$, $p = 0.17$, despite non-mil. (M = 58.86, SD = 13.25) attaining higher scores than mil. (M = 49.83, SD = 8.40).

There was no significant effect on the TCT from trial 4, $t(9.02) = 0.77$, $p = 0.46$, despite non-mil. (M = 38.86, SD = 15.64) attaining higher scores than mil. (M = 30.67, SD = 21.48).

There was no significant effect on the TCT from trial 5, $t(8.79) = 0.95$, $p = 0.37$, despite non-mil. (M = 40.43, SD = 18.79) attaining higher scores than mil. (M = 28.00, SD = 26.85).

There was no significant effect on the TCT from trial 6, $t(8.81) = -0.95$, $p = 0.37$, despite non-mil. (M = 26.57, SD = 21.78) attaining lower scores than mil. (M = 35.33, SD = 10.27).

There was no significant effect on the TCT from trial 7, $t(5.31) = 0.41$, $p = 0.70$, despite non-mil. (M = 29.29, SD = 3.90) attaining higher scores than mil. (M = 25.83, SD = 20.53).

There was no significant effect on the TCT from trial 8, $t(9.94) = -1.15$, $p = 0.28$, despite non-mil. (M = 19.57, SD = 12.22) attaining lower scores than mil. (M = 28.17, SD = 14.37).

There was no significant effect on the TCT from trial 9, $t(5.86) = -1.41$, $p = 0.21$, despite non-mil. ($M = 26.86$, $SD = 3.34$) attaining lower scores than mil. ($M = 33.17$, $SD = 10.53$).

There was no significant effect on the TCT from trial 10, $t(8.33) = -1.24$, $p = 0.25$, despite non-mil. ($M = 22.71$, $SD = 8.64$) attaining lower scores than mil. ($M = 27.17$, $SD = 3.66$).

9.2 Qualitative data analysis

To render the interview transcripts better suitable for analysis, the transcripts went through closer inspection where unnecessary lines including most lines spoken by the interviewers were filtered out, leaving only the answers relevant to the analysis. The initial codes from the resulting transcripts were generated using ATLAS.ti (2024) and its “AI Coding”. Using “Explore the use of a VR simulator as a learning platform for trauma and emergency care” as the prompt for research goal, ATLAS.ti generated the following categories and accompanying questions to use for the coding:

- Effectiveness of VR Simulator - How effective is a VR simulator as a learning platform for trauma and emergency care?
- Pros and Cons of VR Simulator - What are the advantages and disadvantages of using a VR simulator for training in trauma and emergency care?
- Impact on Knowledge Retention - How does the use of a VR simulator impact the retention of knowledge and skills in trauma and emergency care compared to traditional training methods?
- Immersive Learning Experience - In what ways does the immersive nature of a VR simulator enhance the learning experience in trauma and emergency care training?

9.2.1 Themes

The generated codes from ATLAS.ti were processed by filtering out false positives, combining or removing duplicate codes, and removing codes with fewer than four occurrences. This resulted in 34 codes from which the following four themes emerged: *Involvement*, *Training*, *User Experience*, *Accuracy*.

Involvement:

Involvement refers to how engulfed the user was in the scenario and how present they were in the virtual environment.

The participants had rather differing opinions on the level of involvement of the simulator. Some participants expressed that they felt exactly as if they were right there in the scenario or the environment while others always felt very aware that it was not real.

> *You were really into it. It was really reality-based, it felt like. It feels like you were really there and for me who hasn't been in VR before or in any way has done much at all with it, it was a notch higher than I was used to in the artificial (P12, mil.).*

> *It felt like you were really there [...] [you] kind of formed a connection to this person you were supposed to help in some way, more than I expected it to be so my experience was that I was really there (P14, non-mil.).*

> *It is only a computer game, that is, there is no difference no one is dying [...] personally I was not stressed at all because I know that it is a simulation, it is a computer game (P11, non-mil.).*

When asked to recall specific details that the participants noticed throughout the simulation, they mentioned a multitude of aspects. A commonly mentioned aspect was how the physics worked and how the participant perceived the interactions with objects. Some participants thought that the way objects spawned in the world felt unrealistic.

> *It feels a bit artificial how things just sort of popped up next to me and I would just sort of pick up and drop. I was a bit surprised by that (P1, mil.).*

> *Suddenly there was a walkie-talkie next to me. It took me by surprise and I knew I had to report it. It was not something I could have foreseen (P10, non-mil.).*

Two participants mentioned that the rocks did not behave as expected when throwing them which made them feel not as present. One of those participants described the feeling of gravity as extraterrestrial despite the simulator emulating the gravity of Earth.

> *The stones were not at all in gravity as I am used to. It was as if [I] were on another planet (P11, non-mil.).*

> *I think the biggest difference felt like when you threw the stones they went a little 'off' from where I would expect if I threw a stone in real life (P9, non-mil.).*

Two participants with military backgrounds commented on the behaviour of the soldier they were trying to save. They believed the soldier should provide feedback to the user, either by showing signs of life such as breathing or by making physical resistance to indicate pain and thereby evoke stress in the user.

> *You can certainly make it so that the person is not just lying stone dead. Well, hopefully he wasn't dead, but maybe breathing could have an impact in some way (P12, mil.).*

> *It's not the kind of time pressure that puts psychological pressure on, so one feels 'I really have to hurry up'. If [the wounded soldier] had been*

> *screaming and shouting or jumping around or something so that one really felt that this guy needed help now, one would really have felt that 'I had to [help them]' (P13, mil.).*

Training:

Training refers to the simulator's capacity to be used as a training platform for casualty care.

The participants had a lot of opinions on the simulator's capacity for training, in varying levels of abstraction. Some participants commented on it in a very concrete and practical way, such as "Can the simulator be used for training as it exists right now?" or how if a certain feature or aspect of the scenario were implemented differently, the simulator could be used for training.

> *I guess [the simulator] would have to be more complicated. Maybe you could put this tourniquet on like... like pull it up and down instead of just having two positions on it. You could have a little step where you tighten it maybe and have to twist it. Maybe make the mines a little more difficult to see so that you actually need to think about it (P1, mil.).*

> *Yes, I definitely think that it could be used for training. The biggest problem with this one specifically was that it was a little too easy to see the mines and ignore them. I assume that in reality it should not be that easy (P9, non-mil.).*

> *Although you can maybe practice that at home as well or like without a scenario but you can easily have it so that there is a little bit more you also need to do, other stuff. [...] while you're doing this you need to, maybe I don't know like, shoot (P10, non-mil.).*

There were also those who thought about it in a more general and abstract sense, such as "How could this type of simulator be used?" Out of those, most were very optimistic about the simulator's potential to be used for training, identifying potential areas of use.

> *[I] think there can be a lot of different areas where you can use it, [...] fire training kind of how to act in different scenarios [...] but health care was probably the first thing I thought of. Health care and things like first aid (P14, non-mil.).*

> *[...] that is in the training courses I have attended, it is in the armed forces, the police, rescue services, sea rescue, mountain rescue, linked to medical care (P5, mil.).*

> *[...] introduce the sequence of events and what steps and such, for example to children, in case of a fire, the steps of what to do, putting a blanket on it depending on the kind of fire. For those kinds of things I think it could be useful (P8, non-mil.).*

However, for casualty care specifically, the responses were a lot more reserved, especially by the participants with military experience, often pointing out its flaws and identifying that it could at most be used as a complement to regular training.

> *Then, of course, you have to learn the fine motor skills, with your fingers and so on. But that's something you can learn in a classroom or if you lie down and do it practically. But here you can get the scenario thinking in a completely different way. [...] It is not the actual craft, how to put on a tourniquet, which strap should go where, which fingers you should use to tighten [the tourniquet], that is not so damn difficult to teach. But it's more of the bigger picture: when should I do what? What is the right decision to make, well, like here, if someone has stepped on a mine? (P5, mil.).*

> *[...] the application is a bit too easy and it's... the details when you apply this stuff, that's what needs to be practiced, that's what's coming. All this fiddling with pulling it out and applying it right with velcro and twisting and so on, I think you miss too much of what's important if you make it too easy, I would say (P4, mil.).*

> *I wouldn't trust that someone who has perfected this scenario can apply a tourniquet well, but I think it's a good complement (P6, mil.).*

User Experience:

User Experience refers to the user's interaction with the simulator and its implemented functionality in terms of usability and ease of use.

For many participants, the user test was their first-ever experience using a VR system which rendered the interviews an excellent opportunity to obtain their initial reactions and sentiments to the technology. Participants who used VR for the first time in the user tests generally recounted their experience as positive and fun, some expressing surprise at how quickly they became habituated to using the VR system despite never having used anything similar. A few participants also connected their experience in the simulator with playing video games.

> *I've never worked with VR, I've never even looked in [a VR system], so I'm impressed with how well it works and how adaptive you are, how quickly you learn (P5, mil.).*

> *I think the experience has been good. It is clear from the start what is going to happen (P4, mil.).*

> *I found it very interesting. [I] haven't done a lot of VR, so it was fun (P9, non-mil.).*

> *[The experience was] fun, more than anything. Because it felt more like a game it was clear that I have a goal, I want to do this and I want to do it as well as possible (P8, non-mil.).*

Most participants found the controls easy to use and mentioned that they adapted to them without much effort even though they did not feel completely natural.

> *Since I'm so used to playing a lot of games, [the controls] come quite naturally to me. The movements themselves were not unnatural even though they may not be completely realistic, they are still natural for muscle memory (P8, non-mil.).*

> *I feel that I quickly learnt the controls after you went through how to use them and the more attempts I made, the better I got into them and how to use them (P12, mil.).*

Other participants did not feel like they adapted as effortlessly since they were not accustomed to the placement of the buttons on the controllers, but recognised that they would eventually master the controls if they continued using the system regularly.

> *I thought I had good control of the equipment. But then there were a few times I felt like I was pushing the wrong buttons, so I still think I was in control but not to the level that I would like to be. But I also think that if you had done it for a while longer, or if this was a regular thing that you were trained in, it would definitely become natural (P9, non-mil.).*

One participant who was completely new to VR mentioned confusion as one of their first reactions but also acknowledged that it may have been an intentional goal to evoke that feeling for the user.

> *I thought that at first there was a lot of confusion, but maybe that was the point, because you are not really used to or familiar with the environment (P10, non-mil.).*

When commenting on the ease of use of VR, one participant did not believe that the system was difficult to use but could understand why a newcomer to VR would have issues running the simulator without help.

> *I think that just because of VR as a tool, I have a hard time expecting that anyone can just put [the HMD] on and go. You still need some guidance as a first-time user (P7, non-mil.).*

Accuracy:

Accuracy relates to how well the VR simulator corresponded with the participant's real-world experience or expectations. The perceived accuracy of the simulator differed notably between the participants with military and non-military backgrounds.

The main subject of criticism of the scenario was the application of the tourniquet to the soldier. This stage of the scenario was commonly described as simplified, confusing, or misrepresentative of how the process would unfold in a real-life scenario:

> *I don't think the carry-over is so great from clicking a button and picking*

> *up a tourniquet and dropping it near something as in reality it's a bit more of a process that you have to sort of search a person and so on (P1, mil.).*

> *The fact that you just put [the tourniquet] on, I think it would be good if you get behavioural patterns as well [for example] that you have to unfold it properly (P4, mil.).*

> *As I have much more experience in the actual craft of using tourniquets, I'm a bit focused on [the details of the tourniquet application], like which way it should sit, and which flap I should pull. But that is a limitation of [the] human factor and [a VR system] (P5, mil.).*

Some of the participants with military backgrounds noticed that the placement of the tourniquet that the scenario taught differed from what they had been taught throughout their military training.

> *I remarked that the first two times, I set the tourniquet high because that is what we have learnt [in the military] (P5, mil.).*

A few participants without military backgrounds commented that some parts of the scenario felt unnatural. However, the participants did not elaborate further since they believed that their previous experience with tourniquet application was too limited to draw any conclusions on the accuracy of the scenario.

> *There are some things [in the scenario] that felt unnatural but also, the fact that I myself have no experience of such a situation [makes it] difficult to connect to something I would experience (P7, non-mil.).*

Another aspect that participants highlighted was how one moved through space in the scenario. The participants mentioned that teleportation felt unrealistic and rendered the overall sense of movement artificial and too easy.

> *It was mostly the movement as I could quickly get to them without moving my legs (P2, non-mil.).*

> *I could just teleport myself [to the soldier]. Now we're probably limited in this room because of [its layout] but it was a bit easy to just teleport there (P12, mil.).*

> *Pointing and teleporting is quite artificial (P10, non-mil.).*

> *[The movement] felt a bit different to walking. You could also walk, so if you had more space, you would really have the opportunity to make [the movement] feel closer to reality (P13, mil.).*

One of the participants with a military background remarked that the model for landmines that the simulator used was inaccurate. They did not elaborate further on the topic, but it constitutes an apparent inaccuracy in the simulation nonetheless.

> *[I] noticed because of my nerdiness that [the mines] are anti-tank mines but that's another matter. (P5, mil.).*

9.3 Design considerations

To answer the research question clearly, a few design considerations were constructed from the qualitative and quantitative data analysis as seen in table 9.4.

Table 9.4: Considerations for Designing a Casualty Care Simulator for Standalone VR Systems.

No.	Consideration
1	The Effect of Learning Measures on Involvement
2	Possible Training Applications
3	Balance between Serious and Engaging Elements
4	Locomotion

1. The Effect of Learning Measures on Involvement

As described in section 6.2.2, in accordance with CLT, the simulator had implemented measures to minimise extraneous cognitive load while optimising germane cognitive load, helping the user learn the material. An example of this is spawning in the walkie-talkie instantly after signing the tourniquet which diverts the user's attention to it as can be seen in figure 7.8.

These implemented measures were successful since every participant finished at least one trial with a full score, see table 9.1. However, some of these measures seemed to have hurt user involvement. This is exemplified in section 9.2.1 under *Involvement* where participant number 10 described the walkie-talkie spawning in beside the soldier as surprising and not something they could foresee. Participant number 1 commented that every object that popped up next to them felt artificial and surprised them.

As stated in section 1.2, the aim of the simulator is to let users learn the material in an environment resembling the real-world context in which the knowledge will be used. This makes involvement a crucial aspect. That in turn means that when designing a casualty care simulator for standalone VR, one needs to be mindful of the effects implemented measures to help with learning may have on involvement.

2. Possible Training Applications

As described in section 6.2.2, the simulator's lack of the sense of touch and its use of hand controllers meant that actions such as applying and signing the tourniquet had to be simplified down to a button press which does not resemble how the process would be performed in practice. This was commented on by participants number 4, 5, and 6 in the interviews in section 9.2.1 under *Training*, discussing how the simulator cannot entirely replace traditional training methods but could potentially be used as a complement. Specifically, participant number 5 commented that while the simulator cannot teach the minute details of the tourniquet application, it can teach the sequence of actions to take.

To summarise, when designing a casualty care simulator for standalone VR systems, one should consider the strengths and limitations inherent to such systems and how they affect the simulator's possible training applications.

3. Balance between Serious and Engaging Elements

As discussed in section 2.1, a casualty care simulator is developed with the primary goal of educating users in real-life skills that can be put into practice at some point. This stands in contrast to other software which does not need to prioritise realism and inducing stress to the same degree (Binsch et al., 2021). Under *Involvement* in section 9.2.1, participant number 11 commented that the simulator was "only a computer game" and that "I was not stressed at all because I know that it is a simulation". In another interview, under *User Experience* in section 9.2.1, participant number 8 commented that "Since I'm so used to playing a lot of games, [the controls] come quite naturally to me". They further commented that "[The experience was] fun, more than anything. Because it felt more like a game it was clear that I have a goal, I want to do this and I want to do it as well as possible" Furthermore, the RTLX score from section 9.1 reinforces that the experience in the scenario was generally not stressful to the participants.

While it is beneficial to users to be able to connect their experience to playing a game and consider it fun, this should not detract from the seriousness and stress associated with the scenario in order to prevent a disconnect between the simulation and the real-life scenario. As described in section 6.2.1, the scenario should replicate the real-life steps of applying a tourniquet, allowing them to be employed in a real-life emergency. According to Binsch et al. (2021), this also entails inducing other external factors that may arise in practice, most importantly stress. Participant number 13 in section 9.2 under *Involvement* further highlighted the importance of stress in the scenario, stating "It's not the kind of time pressure that puts psychological pressure on, so one feels 'I really have to hurry up'. If [the wounded soldier] had been screaming and shouting or jumping around or something so that one really felt that this guy needed help now, one would really have felt that 'I had to [help them]'".

In conclusion, one should consider how to strike a balance between realistic elements such as stress, and maintaining engaging elements. This balance could help the simulator to remain captivating and effective in preparing users for real-life scenarios.

4. Locomotion

As described in section 2.4 and indicated by its definition, standalone VR systems do not suffer from being constrained by external hardware be it motion trackers or a computer. The scenario as described in section 6.2.1 necessitated the implementation of movement. Therefore, making use of this freedom of movement afforded by standalone VR was an essential aspect of designing the simulator. However, given that the size of the physical space in which the simulator might be used is unknown, some form of virtual locomotion was also necessary in addition to real-world movement. As described in section 6.2.2, teleportation was used as the virtual locomotion technique for the simulator. Under *Accuracy* in section 9.2.1 participants mention how the teleportation made the movement feel artificial and too easy. Participants number 2 and 12 said that they could quickly get to the wounded squadmate without moving in the real world. In other words, the implementation of virtual locomotion had a diminishing effect on the use of real-world movement. Therefore, to fully exploit the advantages of standalone VR, one should consider how virtual locomotion and real-world movement affect one another so that a combination between them can function well as a compromise between natural movement and the practical limitations of physical space.

10

Discussion

This chapter discusses the findings, experiences, and challenges with developing the prototype, performing the user tests, and formulating the design considerations. Also, some ethical considerations and limitations of the study are brought up along with possibilities for future work with the prototype and future user tests.

10.1 Prototype

Since a pilot study was conducted after finishing the first version of the prototype, the process can be described as an iterative design process. Because of the limited time frame for development, only two iterations were completed with the second being considerably shorter than the first. The reason for this was that the requirements for the simulator had already been met after the wrap-up of the first iteration. It is also important to mention that the requirements for the two iterations differed. The requirements for the first iteration came from the provided scenario description while the requirements for the second iteration were solely based on feedback from the pilot studies. Nonetheless, iterative design generally allows for such flexibility when defining such requirements.

Having clear requirements from the beginning facilitated the first stages of development since the process otherwise would include three separate data collection opportunities which would have constrained the development time considerably. This could have been fatal to the quality of the prototype as it was built from scratch in a game engine with which the researchers had little experience. Thus, the clear initial requirements enabled focusing efforts towards a functional and effective prototype without being bogged down by continuous data collection and requirement adjustments. However, the requirements for the first iteration did not explicitly mention any points related to user experience which quickly became apparent during the pilot studies with complaints about unclear and confusing interactions being common. Consequently, the second iteration funnelled efforts towards incorporating feedback from the pilot studies which generally addressed specific user concerns and improved the overall user experience.

More resources on constructing user interactions in VR might have been found with additional research, but time constraints necessitated moving forward. Developing the user experience without a comprehensive framework thus became challenging at times as it made decision-making difficult whenever several valid options existed for converting a multiple-choice action into a pure VR interaction.

The possibility of integrating hand tracking without the controllers was considered during the early stages of prototyping. This could render the interaction and experience in the simulator even more realistic since many participants brought up the controllers as a distracting part of the experience. As previously mentioned in section 2.3, the Meta Quest 3 has built-in support for hand tracking and it could therefore serve as a feasible interaction solution. However, at the time of development, Unreal Engine lacked support for hand-tracking interactions, rendering it difficult to implement. Moreover, there was a noticeable delay between moving the hands and receiving feedback in the HMD. It was for these two reasons that this feature was abandoned early on during the prototyping process. Nevertheless, hand tracking using cameras is still in its early days of development and might become a more prevalent solution in the future as the technology progresses.

10.2 Findings

RQ: *What should be considered when designing a casualty care simulator for standalone VR systems?*

The thesis work resulted in four design considerations based on the quantitative and qualitative data analysis in combination with the literature and theory from chapters 2 and 3 respectively. We realise that some design considerations are rather general and could apply to any VR system that can be used for a casualty care simulator. The two glaring examples here are design considerations number 1, *The Effect of Learning Measures on Involvement*, and number 3, *Balance between Serious and Engaging Elements*, However, these considerations are not less meaningful or useful just because they are general and further help to solidify the principles that govern effective VR design. By addressing general considerations in combination with more specific ones, it could hopefully inspire future casualty care simulators to provide reliable and intuitive user experiences whilst also utilising the strengths specific to standalone VR.

We believe that quantitative data analysis took valuable time that could have been spent on an extended qualitative data analysis. A considerable amount of time was directed toward gathering and analysing quantitative data. First, we had to go through the video recordings for every participant and convert data for every stimulus into a format suitable for analysis. After that, we analysed all of the 28 DVs using quantitative data analysis methods which made for quite a rigorous result. However, none of the 28 DVs turned out to have any statistical significance and we could therefore not use any of them to draw any conclusions from quantitative data alone. Even in the scenario that the results were statistically significant, we do not think it would help us answer the research question. In a best-case scenario, we hoped that the quantitative data analysis would reveal something of value that could aid the qualitative data analysis. In other words, the goal was primarily exploratory and formative more than to perform a summative evaluation of the prototype.

Nonetheless, we believe that a quantitative data analysis could be a better fit in a comparative study between different VR systems. In our study, we used a subject variable which is an inherent characteristic of the participants themselves, limiting the conclusions we could draw from the data analysis. In a comparative study, the VR system would be the IV which would eliminate the inherent limitations of using a subject variable, rendering the result more rigorous.

10.3 Ethical considerations

The portrayal of trauma scenarios can exert psychological strain on users and can potentially lead to stress or trauma-related responses, particularly given the immersive capabilities of VR technology. This aspect underscores the necessity for considering the mental health impacts associated with the use of such simulators in training environments. Therefore, a disclaimer about the inclusion of content that could be perceived as graphic was included in the consent form that all participants had to thoroughly read and sign before taking part in the study.

Another ethical concern is the handling of personal data collected from the study. The study itself was conducted in a closed room with no risk of outside observation, ensuring the privacy of the participants. It was clarified in the consent form that any and all data collected would be completely anonymised. Finally, all collected anonymised data was systematically deleted after the data analysis was finished. This step was taken to ensure that no residual data remained that could potentially compromise the participants' anonymity or be misused in any manner.

Furthermore, there is a risk that casualty care training could become overly reliant on VR technology in the future. This reliance might lead to a reduced emphasis on hands-on, practical training, which is crucial for developing the motor skills necessary to mount a tourniquet tightly and efficiently. However, as discussed in the interviews, most participants stated that the lack of tactility renders VR an unfit replacement for traditional training. Instead, VR training could be integrated with traditional methods to provide a comprehensive training experience that includes both cognitive and physical skill development.

Currently, there are no extensive long-term studies examining the prolonged effects of casualty care VR simulators, leaving a significant gap in the understanding of their impact. As this technology continues to evolve and become more integrated into training programs, ongoing research and longitudinal studies will be essential to fully comprehend the ethical implications and to ensure that the benefits of VR training are maximised whilst minimising any adverse effects. Thus, only time will reveal the true extent of these ethical considerations discussed in this section.

10.4 Limitations

One limitation of the study is the number of participants (n = 13). This could be one reason why the results showed that military experience has no significant effect on any of the DVs. The main contribution and hence the most time-consuming part of the thesis is the simulator. For this reason, there was no time for a study of greater size to be conducted. This means that the study results should be considered part of a formative assessment rather than a summative assessment of the simulator. In other words, this thesis work warrants further iterations in the RtD process to delve deeper and more rigorously answer the research question.

Another limitation is the use of non-probability sampling for recruiting the participants for the user tests which is prone to inducing biases into the sampling process. The process of recruiting participants through probability sampling is also more rigorous since the samples are usually selected from a larger pool of possible participants. Had the study employed probability sampling methods, the findings would likely have been more trustworthy, providing stronger evidence and more accurate insights into the research question.

Rendering the analysis even more difficult was the fact the presence questionnaire did not have a standardised way of calculating and interpreting scores. For this reason, the scores from the presence questionnaire could only be used to compare the participants within the study and not to compare with previous work.

10.5 Future work

Future work should involve further simulator development, focusing on adding more variance to the scenario. This could include small changes like altering which foot gets amputated or changing the position of the squadmate in the world, but also larger changes such as setting the scenario in different environments. By introducing this variance, users can better learn the procedure and avoid becoming overly reliant on concrete implementations.

Another improvement to the scenario is to add more sounds to the environment. In its current state, the scenario has a few sounds such as background nature sounds, explosion sounds when a mine explodes, and sounds for signing and tightening the tourniquet. However, some participants mentioned that they thought some areas of the scenario could benefit from more auditory feedback. For example, it was unclear to some participants whether the squadmate was dead or not. Having the squadmate give breathing sounds or screams could therefore help clarify to the user that he was indeed still alive.

Moreover, casualty care involves more than just applying a tourniquet, especially when considering applications beyond the military context. Interviews revealed potential uses in fire safety training and search and rescue scenarios. Therefore, adding more unique scenarios would not only expand the simulator's utility but also help in its evaluation and in building a broader framework for casualty care simulators on standalone VR systems.

Additionally, conducting a more comprehensive study with a larger sample size selected through probabilistic sampling could yield more substantial and deeper insights into the context and implementation requirements for standalone VR casualty care simulators. Furthermore, a comparative study between different casualty care training platforms including traditional methods and other digital platforms could be beneficial. As mentioned in section 2.4, Tactisim is developing a platform that combines VR with real elements, providing users with physical aspects to interact with. Such a study could help identify the strengths and weaknesses of different training approaches and platforms, guiding future development efforts effectively.

As previously mentioned in section 6.1.1, hand tracking could become a viable option for interaction in place of physical controllers which the user must hold in their hands. Hand tracking would open up a flood of new interaction functionality that could transform the simulator considerably since all interactions could be made more realistic and natural to users. It would also allow users to interact with real-life objects that simultaneously get rendered virtually in the simulator, making the experience even more realistic.

11

Conclusion

The thesis project aimed to develop a prototype of a casualty care simulator for the Meta Quest 3. The prototype was developed around a scenario in which a squadmate triggers a landmine and amputates their foot, forcing the user to take action and save them as quickly as possible. The purpose of developing the prototype was to answer the research question *What should be considered when designing a casualty care simulator for standalone VR systems?* To evaluate the prototype, a study was conducted using participants (n = 13) with and without military experience. The study collected both quantitative and qualitative data that was deemed relevant to help answer the research question. The following statistical and thematic analysis of the quantitative and qualitative data revealed important insights that allowed the research question to be answered.

The quantitative data analysis found that although the scenario followed a sequence of events that people with military experience would know rather well, there was no correlation between performance and military experience. The interviews and subsequent analysis of the collected data in the study culminated in four design considerations for designing a casualty care simulator for standalone VR. An important takeaway highlighted by design consideration 2 is that standalone VR might not be able to entirely replace traditional training methods but instead be suitable as a complement.

Given the limited scope of the thesis and its study, this thesis warrants future work as outlined in the discussion.

Bibliography

Aggelidis, V. P., & Chatzoglou, P. (2009). Using a modified technology acceptance model in hospitals. *International journal of medical informatics*, *78 2*, 115–26. https://doi.org/10.1016/j.ijmedinf.2008.06.006

Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2024, February 15). Hypothesis testing [Accessed: 2024-02-19]. https://www.britannica.com/science/statistics/Hypothesis-testing

ATLAS.ti. (2024). https://atlasti.com/

Baker, R., Brick, J., Bates, N., Battaglia, M., Couper, M., Dever, J. A., Gile, K., & Tourangeau, R. (2013). Summary report of the aapor task force on non-probability sampling. *Journal of Survey Statistics and Methodology*, *1*, 90–143. https://doi.org/10.1093/JSSAM/SMT008

Barteit, S., Lanfermann, L., Bärnighausen, T., Neuhann, F., & Beiersmann, C. (2021). Augmented, mixed, and virtual reality-based head-mounted devices for medical education: Systematic review. *JMIR Serious Games*, *9*(3), e29080. https://doi.org/10.2196/29080

Benyon, D. (2010, April). *Designing interactive systems* (2nd ed.). Addison-Wesley Educational.

Berkenstadt, H., Ziv, A., Barsuk, D., Levine, I., Cohen, A., & Vardi, A. (2003). The use of advanced simulation in the training of anesthesiologists to treat chemical warfare casualties. *Anesthesia & Analgesia*, *96*, 1739–1742. https://doi.org/10.1213/01.ANE.0000057027.52664.0B

Binsch, O., Bottenheft, C., Landman, A., Roijendijk, L., & Vermetten, E. H. (2021). Testing the applicability of a virtual reality simulation platform for stress training of first responders. *Military Psychology*, *33*(3), 182–196.

Bond, D., & Nyblom, M. (2019). Evaluation of four different virtual locomotion techniques in an interactive environment.

Broeck, J. v. d., & Brestoff, J. R. (2013). *Epidemiology: Principles and practical guidelines*. Springer.

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

Butler, F. (2017). Leadership lessons learned in tactical combat casualty care. *The journal of trauma and acute care surgery*, *82 6S Suppl 1*, S16–S25. https://doi.org/10.1097/TA.0000000000001424

Butler, F. K., & Kotwal, R. S. (2017). Tactical combat casualty care. *Front Line Surgery: A Practical Approach*, 3–16.

Commons, C. (2024). https://creativecommons.org/share-your-work/cclicenses/

Coomer, N., Bullard, S., Clinton, W., & Williams-Sanders, B. (2018). Evaluating the effects of four vr locomotion methods: Joystick, arm-cycling, point-tugging, and teleporting. *Proceedings of the 15th ACM symposium on applied perception*, 1–8.

Dam, R. F., & Teo, Y. S. (2024). 5 common low-fidelity prototypes and their best practices. *Interaction Design Foundation - IxDF*. https://www.interaction-design.org/literature/article/prototyping-learn-eight-common-methods-and-best-practices

Davis, F. D. (1985). *A technology acceptance model for empirically testing new end-user information systems: Theory and results* [Doctoral dissertation, Massachusetts Institute of Technology].

Emerson, R. W. (2015). Convenience sampling, random sampling, and snowball sampling: How does sampling affect the validity of research? *Journal of visual impairment & blindness*, *109*(2), 164–168.

Fiani, B., Stefano, F. D. D., Kondilis, A., Covarrubias, C., Reier, L., & Sarhadi, K. (2020). Virtual reality in neurosurgery: "can you see it?" a review of the current applications and future potential. *World neurosurgery*. https://doi.org/10.1016/j.wneu.2020.06.066

Geyer, F., Pfeil, U., Höchtl, A., Budzinski, J., & Reiterer, H. (2011). Designing reality-based interfaces for creative group work, 165–174. https://doi.org/10.1145/2069618.2069647

Gillett, B., Peckler, B., Sinert, R., Onkst, C., Nabors, S., Issley, S., Maguire, C., Galwankarm, S., & Arquilla, B. (2008). Simulation in a disaster drill: Comparison of high-fidelity simulators versus trained actors. *Academic Emergency Medicine*, *15*(11), 1144–1151.

Government of Canada, S. C. (2021, September). 3.2.2 Probability sampling [Last Modified: 2021-09-02]. Retrieved May 17, 2024, from https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch13/prob/5214899-eng.htm

Harboe, G., & Huang, E. (2015). Real-world affinity diagramming practices: Bridging the paper-digital gap. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/2702123.2702561

Hart, S. G. (2006). Nasa-task load index (nasa-tlx); 20 years later. *Proceedings of the human factors and ergonomics society annual meeting*, *50*(9), 904–908.

Heinrichs, W., Youngblood, P., Harter, P. M., & Dev, P. (2008). Simulation for team training and assessment: Case studies of online training with virtual worlds. *World Journal of Surgery*, *32*, 161–170. https://doi.org/10.1007/s00268-007-9354-2

Hertzum, M. (2021). Reference values and subscale patterns for the task load index (tlx): A meta-analytic review. *Ergonomics*, *64*(7), 869–878. https://doi.org/10.1080/00140139.2021.1876927

HTC. (2024). Vive pro 2 headset [Accessed: 24-05-2024]. https://www.vive.com/eu/product/vive-pro2/specs/

Kahn, M. J., Tan, K. C., & Beaton, R. J. (1990). Reduction of Cognitive Workload through Information Chunking. *Proceedings of the Human Factors Society Annual Meeting*, *34*(19), 1509–1513. https://doi.org/10.1177/154193129003401919

Kaptelinin, V. (2014). Activity theory. Interaction Design Foundation - IxDF. https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/activity-theory

Karanasios, S., Nardi, B., Spinuzzi, C., & Malaurent, J. (2021). Moving forward with activity theory in a digital world. *Mind, Culture, and Activity*, *28*(3), 234–253.

Kleygrewe, L., Hutter, R. V., Koedijk, M., & Oudejans, R. R. (2023). Virtual reality training for police officers: A comparison of training responses in vr and real-life training. *Police Practice and Research*, *25*(1), 18–37.

Lovreglio, R., Duan, X., Rahouti, A., Phipps, R., & Nilsson, D. (2021). Comparing the effectiveness of fire extinguisher virtual reality and video training. *Virtual Reality*, *25*(1), 133–145. https://doi.org/10.1007/s10055-020-00447-5

Luchs, M. G., Swan, S., & Griffin, A. (2015). *Design thinking: New product development essentials from the pdma*. John Wiley & Sons.

MacKenzie, I. S. (2013). *Human-computer interaction: An empirical research perspective* (1st). Morgan Kaufmann Publishers Inc.

Marín-Morales, J., Llinares, C., Guixeres, J., & Alcañiz, M. (2020). Emotion recognition in immersive virtual reality: From statistics to affective computing. *Sensors*, *20*(18), 5163.

Marshall, M. (1996). Sampling for qualitative research. *Family practice*, *13 6*, 522–5. https://doi.org/10.1093/FAMPRA/13.6.522

McAvinia, C. (2016). *Online learning and its users: Lessons for higher education* [OCLC: 948247505]. Chandos Publishing, an imprint of Elsevier.

Mehrfard, A., Fotouhi, J., Taylor, G., Forster, T., Armand, M., Navab, N., & Fuerst, B. (2021). Virtual reality technologies for clinical education: Evaluation metrics and comparative analysis. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, *9*(3), 233–242. https://doi.org/10.1080/21681163.2020.1835559

Merriënboer, J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, *17*, 147–177. https://doi.org/10.1007/S10648-005-3951-0

Meta. (2024a). Full-colour pass-through on meta quest [Accessed: 21-05-2024]. https://www.meta.com/en-gb/help/quest/articles/getting-started/getting-started-with-quest-pro/full-color-passthrough/

Meta. (2024b). Meta quest 3 [Accessed: 20-05-2024]. https://www.meta.com/se/en/quest/quest-3/

Naderifar, M., Goli, H., & Ghaljaie, F. (2017). Snowball sampling: A purposeful method of sampling in qualitative research. *14*. https://doi.org/10.5812/SDME.67670

Nardi, B. A. (1995, November). Studying Context: A Comparison of Activity Theory, Situated Action Models, and Distributed Cognition. In B. A. Nardi (Ed.), *Context and Consciousness* (pp. 69–102). The MIT Press. https://doi.org/10.7551/mitpress/2137.003.0008

Nikolopoulou, K. (2022a, August 11). *What is purposive sampling? | definition & examples* [Scribbr]. Retrieved May 15, 2024, from https://www.scribbr.com/methodology/purposive-sampling/

Nikolopoulou, K. (2022b, August 12). *What is quota sampling? | definition & examples* [Scribbr]. Retrieved May 15, 2024, from https://www.scribbr.com/methodology/quota-sampling/

O'Brien, H. L., Cairns, P., & Hall, M. (2018). A practical approach to measuring user engagement with the refined user engagement scale (ues) and new ues short form. *International Journal of Human-Computer Studies*, *112*, 28–39.

Operative Experience. (2024). A global leader in healthcare simulation & tactical medicine [Accessed: 2024-02-11]. https://operativeexperience.com/

Ortiz, M., Sarabia, L., & Sánchez, M. (2010). Tutorial on evaluation of type i and type ii errors in chemical analyses: From the analytical detection to authentication of products and process control. *Analytica Chimica Acta*, *674*(2), 123–142. https://doi.org/https://doi.org/10.1016/j.aca.2010.06.026

Pasquier, P., Dubost, C., Boutonnet, M., Chrisment, A., Villevieille, T., Batjom, E., Bordier, E., Ausset, S., Puidupin, M., Martinez, J., Bay, C., Escarment, J., Pons, F., Lenoir, B., & Mérat, S. (2014). Predeployment training for forward medicalisation in a combat zone: The specific policy of the french military health service. *Injury*, *45 9*, 1307–11. https://doi.org/10.1016/j.injury.2014.05.037

Peckler, B., Gillett, B., Nabors, S., Maguire, C., Issley, S., Onkst, C., Arquilla, B., & Galwankar, S. (2007). 41: Simulation in a disaster drill: Comparison of high-fidelity simulators vs trained actors. *Annals of Emergency Medicine*, *50*. https://doi.org/10.1016/J.ANNEMERGMED.2007.06.072

Pottle, J. (2019). Virtual reality and the transformation of medical education. *Future Healthcare Journal*, *6*(3), 181–185. https://doi.org/10.7861/fhj.2019-0036

Quixel. (2024). https://quixel.com/bridge

Scalese, R. J., Issenberg, S., Hackett, M., Rodriguez, R. D., Brotons, A. A., Gonzalez, M., Geracci, J., & Schulman, C. (2022). Simulation-based education improves military trainees' skill performance and self-confidence in tourniquet placement: A randomized controlled trial. *Journal of Trauma and Acute Care Surgery*, *93*, S56–S63. https://doi.org/10.1097/TA.0000000000003702

Schulz, C., Skrzypczak, M., Raith, S., Hinzmann, D., Krautheim, V., Heuser, F., Mayer, V., Kreuzer, C., Himsl, M., Holl, M., Lipp, C., Kochs, E., & Wagner, K. (2014). High-fidelity human patient simulators compared with human actors in an unannounced mass-casualty exercise. *Prehospital and Disaster Medicine*, *29*, 176–182. https://doi.org/10.1017/S1049023X14000223

Schuster, D. P., & Powers, W. J. (2005). *Translational and experimental clinical research*. Lippincott Williams & Wilkins.

Sharp, H., Preece, J., & Rogers, Y. (2019). *Interaction design: Beyond human-computer interaction* (Fifth edition). Wiley.
Literaturverzeichnis: Seite 581-618. - Index.

Sketchfab. (2024). https://sketchfab.com/

Soegaard, M. (2023). System usability scale for data-driven ux. *Interaction Design Foundation - IxDF*. https://www.interaction-design.org/literature/article/system-usability-scale

Street Smarts VR. (2021). Train like you're there, respond at your best [Accessed: 2024-02-13]. https://www.streetsmartsvr.com/

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, *12*(2), 257–285.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Tactisim. (2024). Saving lives through quality training [Accessed: 2024-02-11]. https://www.tactisim.com/

Thompson, S. K. (2012). *Sampling* (3rd ed) [OCLC: ocn746489136]. Wiley.

University of Oxford. (2021, December 2). Informed consent [Accessed: 2024-02-19]. https://researchsupport.admin.ox.ac.uk/governance/ethics/resources/consent

Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Using presence questionnaires in reality. *Presence*, *9*(5), 497–503.

van Peer, W., Hakemulder, F., & Zyngier, S. (2012). *Scientific methods for the humanities*. John Benjamins. https://doi.org/10.1075/lal.13

Wang, C.-C., Lo, S.-k., & Fang, W. (2008). Extending the technology acceptance model to mobile telecommunication innovation: The existence of network externalities. *Journal of Consumer Behaviour*, *7*, 101–110. https://doi.org/10.1002/CB.240

Williamson, J. H. (2022, August). An Introduction to Bayesian Methods for Interaction Design. In J. H. Williamson, A. Oulasvirta, P. O. Kristensson, & N. Banovic (Eds.), *Bayesian Methods for Interaction and Design* (1st ed., pp. 3–80). Cambridge University Press. https://doi.org/10.1017/9781108874830.003

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence*, *7*(3), 225–240.

Wohlgenannt, I., Simons, A., & Stieglitz, S. (2020). Virtual reality. *Business & Information Systems Engineering*, *62*, 455–461.

Yang, K., & Banamah, A. (2014). Quota sampling as an alternative to probability sampling? an experimental study. *Sociological research online*, *19*(1), 56–66.

Zimmerman, J., Forlizzi, J., & Evenson, S. (2007). Research through design as a method for interaction design research in hci. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 493–502.

Zimmerman, J., Stolterman, E., & Forlizzi, J. (2010). An analysis and critique of research through design: Towards a formalization of a research approach. *proceedings of the 8th ACM conference on designing interactive systems*, 310–319.

A

Flowchart: Scenario

Figure A.1 on the next page depicts the flowchart of the initial scenario provided by Tactisim.

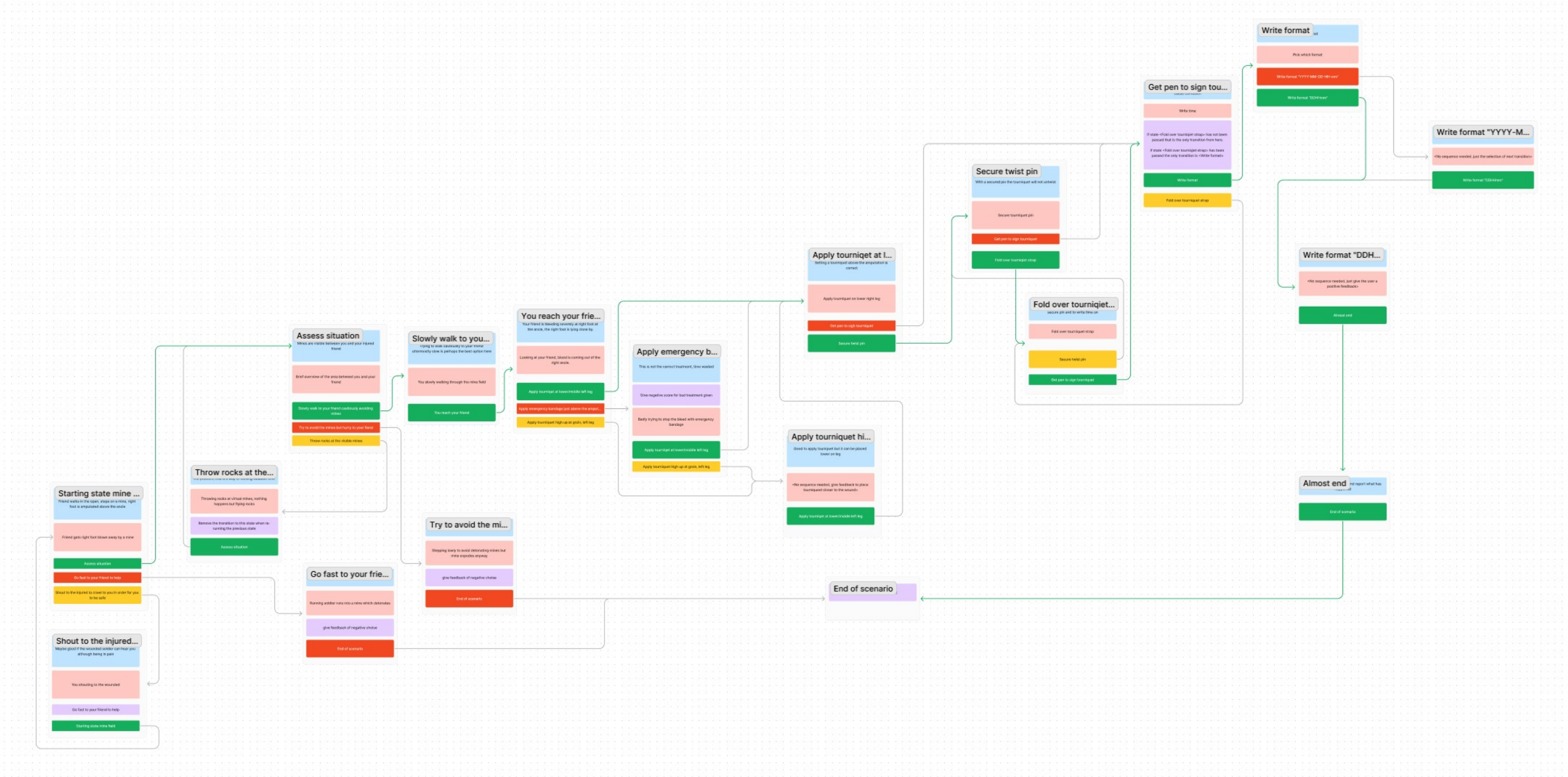


Figure A.1: Flowchart of the scenario provided by Tactisim.

B

Consent form

Welcome to our research study!

We are interested in understanding key aspects to be considered when designing a casualty care simulator for virtual reality (VR). Specifically, we look at user acceptance and user learning. You will take part in a fictitious medical scenario in VR involving a severely injured person where you will have to take the correct and necessary steps to resolve or stabilize the situation. The scenario might be considered graphic since it includes visible blood and an amputated limb. Please also be advised that if you are prone to motion sickness, VR might have a similar effect if used for an extended period of time.

After the scenario in virtual reality you will answer a questionnaire about your experience. This will then be followed up by a short interview.

In order to gather data from the study we will be video recording your real-world actions during the scenario, screen recording your actions in the virtual environment, and audio recording the interview. Please be assured that any and all results gathered from your participation will be completely anonymised.

The study should take you approximately 45 minutes to complete in its entirety. Your participation in this research is voluntary. You have the right to withdraw at any point during the study.

By agreeing to partake in this study, you acknowledge that your participation in the study is voluntary, that you are over 18 years of age, and that you are aware that you may choose to terminate your participation in the study at any time and for any reason.

1. I consent. Begin the study. ________________________

2. I do not consent. Terminate the study. _____________

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

UNIVERSITY OF GOTHENBURG

CHALMERS
UNIVERSITY OF TECHNOLOGY