A FRAMEWORK FOR AUTHENTIC ASSESSMENT IN ROOM-SCALE VIRTUAL REALITY TRAINING SIMULATIONS

SCOTT RONALD HOWIE

Thesis submitted in partial fulfilment of the requirements
of the University of the West of Scotland
for the award of Doctor of Philosophy

April 2021

Abstract

Authentic assessment provides a rigorous method of measuring trainee performance by analysing their application of knowledge for solving real-world problems in a practice environment.

Virtual Reality (VR) training simulations allow training scenarios to be performed remotely in an immersive life-like environment without any danger or safety concerns for the trainee. However, existing methods of conducting authentic assessment and generating feedback of room-scale VR scenarios are inadequate as they are: (i) unable to capture all simulation data; (ii) predefined recordings, meaning trainee performances can be obscured, and (iii) disconnected in their approach in conducting assessment and generating contextual feedback.

This work, grounded in the literature, was developed in consultation with industry training assessors to identify the scope, needs, and requirements of a novel authentic assessment framework for room-scale VR simulations. The findings from industry inspired the design of the components, tools, and features of the framework which enable assessors to conduct authentic assessment and generate feedback for room-scale VR trainee performances.

To evaluate the novel framework an implementation of it was developed. A comparative study involving training assessors from industry and academia compared analytical data and video recording and assessment, along with the implemented framework. Results suggest that the framework satisfied the requirements of assessors, showing positive usefulness and long-term usability, indicating that the framework was the most effective and efficient method for conducting assessment and generating feedback. In comparison, analytical data was found to be unsuitable for authentic assessment due to its lack of context for trainee actions and movement. Whilst, video assessment demonstrated satisfactory usefulness and usability results, it was seen as a passive medium, constrained by its predefined method of capturing and presenting data from a single point-of-view, reducing the interactivity and reliability of assessment and occasionally inducing motion sickness in the assessors.

This thesis original contribution to knowledge lies in the design and a validated implementation of a novel framework for authentic assessment in room-scale VR training simulations, and a novel method for dual-recording of real time VR training, which is patent pending. The novel framework introduced in this work overcomes existing issues of assessment and feedback in VR by: (i) structuring authentic assessment of VR room-scale simulations; (ii) utilising a dual-recording methodology to capture all simulation data with negligible impact on run-time performance, and; (iii) embedding assisted assessment and feedback tools to streamline the evaluation process for assessors.

The originality of this work stems from the design and development of a novel framework, which includes a patent-pending dual-recording methodology. This framework is backed by original contextual design research performed within an industrial training company, which identified the necessary requirements for a framework that would enable assessors to perform authentic assessment in room-scale virtual reality training simulations.

Contents

List of Tables

List of Figures

Listings

Declaration

The research presented in this thesis was carried out by the undersigned. No part of the research has been submitted in support of an application for another degree or qualification at this or any other University.

Acknowledgements

First, I would like to express my sincere gratitude to my Director of Studies, Dr. Marco Gilardi. Your guidance and support was invaluable for completing this project. I have been lucky to have you as a supervisor, knowing that you have always believed in my work and have went above and beyond at every occasion. It would not have been possible to have achieved my ambitions without your continual support and guidance. It has been a pleasure to have you as my supervisor.

I would also like to express my gratitude to all staff at the University of the West of Scotland, but most notably, Prof. Thomas Connolly and Dr. Paul Keir who inspired me to pursue my PhD, and offered their support and guidance throughout my undergraduate and postgraduate studentship at the University of the West of Scotland.

Furthermore, I would like to thank my colleagues at the UWS Immersive lab who were a joy to work beside.

I also lucky to have had the pleasure of working with many great people at Swagelok Scotland ltd. who generously offered their resources, time and assistance for evaluating my research and allowing me to visit their training facility. Most notably I would like to thank Peter O'Connor, Bill Dickie and Sam McCulloch, your generosity and willingness to collaborate with my research was invaluable to the success of my PhD.

Finally, I would like to thank my parents, John and Alison, my sister Rachel, and my dog Summer, who have been a source of inspiration and guidance, their support allowed me to overcome the most difficult times of this project. I would also like to thank the rest of my family and friends for supporting me throughout my research.

Acronyms

VR Virtual Reality

HMD Head Mounted Display

PoV Point of View

FoV Field of View

Chapter 1

Introduction

The recent introduction of Virtual Reality (VR) technologies has made possible to deliver training in ways that were not available a decade ago. However, because of the novelty of this emerging technology methods for assessment and feedback in room-scale VR are not established yet. This work aims to propose an approach to authentic assessment and contextual feedback that will allow assessors to effectively evaluate trainees that have trained in a room-scale VR simulation.

In this chapter the core concepts of this approach for authentic assessment and contextual feedback of room-scale virtual reality training simulations will be introduced. The approach will consist of a novel framework that embeds tools that enable assessors to perform authentic assessment and contextual feedback in room-scale VR. An outline of the proposed framework and the structure of the components is presented, providing an introduction to the content of this work. The chapter concludes by highlighting the author's contribution to knowledge from published and submitted publications.

1.1 Motivation

Authentic assessment is a method of measuring trainee performance by analysing their application of knowledge for solving real-world problems in a practice environment (Wiggins, 1990), which provide natural, authentic and well-constructed means of measuring trainee competency (Boud, 1995). With authentic assessment, trainees are tasked with completing objectives that reflect real-world problems and situations (Gulikers, Bastiaens, and Kirschner, 2004; Gielen, Dochy, and Dierick, 2003), applying their knowledge in a way that is meaningful and viable for real-world operations (Wiggins, 1990), emphasising a hands-on approach to problem solving (Collins, 2002).

World-wide, in-context authentic assessment is conducted in a variety of fields and industries to prepare and improve trainee's knowledge and skill-sets for facing real-world events. However, real-world training sessions are often difficult and costly to schedule and prepare due to inconsistent use of training locations and presentation of training content (Andreatta et al., 2010; Guterman, 2005). For example, in 2004, the German government simulated fourteen different disaster scenarios across the country (Maureen Connolly, 2018). The simulated event took two years to plan and involved the combined efforts of tens of thousands of emergency personnel for a thirty-six hour simulation (Maureen Connolly, 2018). During the simulated scenario, emergency services and personnel were still required to remain prepared to quickly re-deploy to a real-scenario should an emergency arise (Bredl et al., 2015).

Overall, teaching, training and assessment exercises are designed to improve teamwork, confidence, and preparedness (Perry, 2004), and measuring trainee competency (Fardinpour, Reiners, and Wood, 2018; Gulikers, Bastiaens, and Kirschner, 2004). However, real-world simulations often suffer from a wide variety of limitations and constraints, including: cost, equipment and skilled personnel required, scheduling conflicts, inconsistent training contexts and lack of personal feedback (Andreatta et al., 2010; Mossel et al., 2017; Hsu et al., 2013; Ragazzoni et al., 2015; Bucarelli, Zhang, and Wang, 2018). Virtual reality (VR) simulations are touted as a viable alternative to real-world exercises and training (Hsu et al., 2013; Andreatta et al., 2010; Ragazzoni et al., 2015; Vincent et al., 2008), allowing for a consistent testable environment, reduced costs, and reusable scenarios. As such, virtual training simulations have been adopted by many industries, for instance one of the most well-known examples is the aviation industry (VRM Switzerland, 2021) where pilots can train in any flight condition without any safety risks to personnel (McGreevy, 2005). In fact, the European Union Aviation Safety Agency (EASA) has recently approved VR flight simulation as a method for certifying pilots (VRM Switzerland, 2021). With virtual reality head mounted display hardware becoming more economically accessible, virtual training is moving towards an immersive full-body training format, allowing greater immersion and presence compared to conventional desktop systems (Jerald, 2016).

Virtual training simulations can provide high fidelity experiences that imitate the challenges of real-world problems, allowing trainees to demonstrate their skills and application of knowledge in a safe environment (Wood et al., 2013; Vos, 2015). The adoption of VR for training by organisations and industry, suggest there is a growing acceptance of VR to complement real-world training for security (FEMA, 2018; Homeland Security EDGE, 2018), medical (Collins, 2018; Homeland Security, 2018; Virtual Medical Center, 2018), and military organisations (Veterans Affairs, 2018), which inspired to focus this research within the field of virtual reality, rather than augmented reality or serious games. As such, virtual simulations can provide the needed abstraction from the real-world, while still fulfilling the necessary requirements for training, tailoring learning to trainees abilities and needs (Gulikers, Bastiaens, and Kirschner, 2004).

However, the implementation and adoption of VR creates problems for teaching and training, notably with the regards to the difficulty for conducting assessment and generating feedback compared to traditional classroom-based approaches (Hamilton et al., 2021). The practice of providing formative authentic assessment and contextual feedback during education and training are important elements of learning process (Raymond and Usherwood, 2013), which improve knowledge of a subject for the trainee or learner (Black and Wiliam, 1998). One of the main advantages for VR teaching, learning and assessment is the ability deploy a consistent training simulation globally, for remote trainees to experience immersive real-world challenges without risk of harm (Konakondla, Fong, and Schirmer, 2017; Carruth, 2017). However, the virtual performance of a remote trainee is often difficult to determine, with assessors relying on data provided by trainee replies to digital quizzes and tests (Ellaway and Masters, 2008), or from automatic collection of analytical data from simulations. Efforts to observe trainees behaviour within virtual reality applications are being developed (Thanyadit, Punpongsanon, and Pong, 2020; MacCormick et al., 2019). However, these efforts do not provide contextual information regarding the virtual environment in which the trainee is currently located. No framework exists that enables assessors to conduct authentic assessment and generate contextual feedback of room-scale virtual reality training simulation performances.

This work is motivated by the training industry need to close this gap in knowledge by designing and developing a framework that is useful and usable for assessors to conduct authentic assessment and generate contextual feedback. The objective is to enable assessors the ability to monitor, observe, assess and provide feedback of virtual training simulations, through capturing and reconstructing the training experience. Without suitable means to monitor and assess trainee performances in remote applications, the effectiveness of trainee participation and engagement cannot be determined. In automatic assessment of training simulations the outcomes of the assessment are not guaranteed to be accurate (Wijewickrema et al., 2017; Fardinpour and Dreher, 2012), therefore, emphasis is placed on human assessors. Furthermore, as training software becomes more realistic and life-like, the ability to identify trainee problems, mistakes and errors in remote training sessions becomes a significant barrier to the authentic assessment and contextual feedback process, requiring evaluation methods that can identify the root cause and source of trainee errors (Hanoun and Nahavandi, 2018). As such, a framework is required to captures trainee performances in VR simulations and reproduce them in a format that enables assessors to directly observe them, conduct authentic assessment and generate contextual feedback. This design and development of this framework is the drive and motivation of this work. To achieve this, it is important to understand the importance and role of assessment for VR training simulations.

1.1.1 Assessment of Virtual Reality Experiences

A significant body of research on VR development environments, with emphasis on providing developers the tools required to develop VR simulations or applications is available in the literature (Dunne and McDonald, 2010; Takala, 2014; Farra et al., 2016; Kiourt, Koutsoudis, and Pavlidis, 2016; Ritter Iii, Borst, and Chambers, 2015), yet limited research exists within the scope of room-scale virtual reality simulations that focus on how virtual trainee performance can be assessed or reviewed by an assessor, demonstrating a gap in research for support of trainee learning and assessment (Fardinpour and Dreher, 2012). At present, training simulations often rely on Artificial Intelligence (AI) systems to replace human assessors (Lahanas, Georgiou, and Loukas, 2016), missing out on the wealth assessors' experience can provide to trainees. While it is appreciated that automatic assessment using AI is considered an important objective of future research, the outcomes of AI assessment is not yet suitable to replace human assessors (Fardinpour and Dreher, 2012; Wijewickrema et al., 2017). Furthermore, to enable scoring and weighting trainee performance outcomes, AI systems will still require a form of data acquisition that enables them to access vast amounts of simulation data (Upadhyay and Khandelwal, 2019; Mirchi et al., 2020). Subsequently, this research could also contribute to AI research, being relevant for data collection, data presentation and AI training or deep-learning use-cases. As training simulations can include a wide variety of dynamic tasks and objectives, there are too many variables at play for automatic assessment to be considered reliable to replace human assessors.

As such, human focused assessment conducted by assessors is required to support trainee learning in virtual training simulations (Fardinpour, Reiners, and Wood, 2018). The role of human assessors is to "observe and analyse the training" (Fardinpour, Reiners, and Wood, 2018, p. 82), with the overall objective of supporting and correcting the trainees acquisition of knowledge. Left alone without authentic assessment and contextual feedback to support them, the trainee may be unable to timely learn from their mistakes, reducing the opportunity for further learning and future improvement (Deeley, 2018).

In authentic assessment, the performance, procedures and outcomes are all considered part of the assessment, with the simulation used as a vehicle to demonstrate trainee capabilities (Messick, 1994). The ability to capture, replay and review trainee performances is an important tool for conducting authentic assessment (Lazar, Feng, and Hochheiser, 2017; Wiggins, 1990; Gulikers, Bastiaens, and Kirschner, 2006), and provides an assessor with the ability to directly observe trainee performances. The advantage is that assessors are able to archive trainee performances, which can then be used for continued assessment of performance, such as skill decay measurement of surgical accuracy (Mohamadipanah et al., 2020). Currently, to observe trainees using a VR simulation an observer needs to be physically present with the trainee to observe body movements in the real-world, take notes of what the trainee does or video record and screen capture the trainee virtual performance. Although rigorous, this type of observation is difficult, requires large data storage to archive the videos, can impact the simulation performance and offers limited insight of the trainee experience and interactions during the simulation.

Performing contextual observation of trainees in VR is challenging, as the trainee and the assessor are positioned in two different contexts, the assessor is in the real-world whilst the trainee is in the virtual context. Observing the trainee from inside the same VR simulation allows insight into their performance (Hanoun and Nahavandi, 2018) and helps in determining the cause and effect relationships from trainee actions (Hanoun and Nahavandi, 2018). Ideally the trainee or assessor would experience the exact perspective of trainee. Direct observation of trainees in their context can offer insight on design challenges (Lazar, Feng, and Hochheiser, 2017) and circumvents issues often encountered of trainees describing inaccurately what they did due to a lack of awareness or understanding of the task or system under study (Blomberg, Burrell, and Guest, 2007). However, direct perspective observation can induce motion sickness on the observer (Lopez et al., 2017), and the presence of additional recording equipment could cause unnecessary pressure and nervousness (Nyström et al., 2014). These issues are caused by the implementation of existing approaches.

Traditionally, to observe a trainee performance, the observation is conducted from mirrored perspectives of the VR trainees virtual view output onto a 2D-screen. However, this approach does not guarantee knowledge of the state and location of the input devices when the trainee is not looking at them (see Figure 1.1), and is not an accurate representation of the trainee vision, as the field-of-view is not identical. Unlike static desktop training systems, VR equipment has six-degrees of freedom (Tromp, Steed, and Wilson, 2003). Due to VR's unrestricted movement ability, usability issues are common for inexperienced trainees (Tromp, Steed, and Wilson, 2003), with simulation interactions often a combination of movement and input, observation of the trainee posture and controller input is required to identify trainee intentions or problems. Figures 1.1 and 1.2 show the point-of-view differences between first-person perspective observation of the trainee, and the third-person perspective, which offers more context for observation. However, capturing third-person trainee performances through video recording is not viable as room-scale movement prevents a single point-of-view from reliably capturing the trainee performance. Therefore, an alternative method for capturing trainee data for observation and assessment is required. To generate feedback, assessors require the ability to monitor, observe and assess a trainee performance to determine actions or application of logic within the environment (Fardinpour, Reiners, and Wood, 2018).

Figure 1.1: Point-of-View capture of the trainee performance. This is the perspective that is used for video assessment method during the evaluation of the framework (see Section 3.8.1).

1.1.2 Feedback of Virtual Reality Experiences

In real-world training, the direct observation of a participant enables the investigator to determine via visual and audio cues how the trainee is coping with the tasks and objectives (Fardinpour, Reiners, and Wood, 2018). These observations are used to direct a assessors intervention if the

Figure 1.2: Third-person capture of the trainee performance. This is the perspective used for the framework presented in this work.

participant is struggling or is requiring further assistance. By providing feedback the assessor can correct or confirm the actions of a trainee, allowing an assessment of their performance to be ascertained (Shute, 2009). This feedback is used by the trainee to gauge their understanding of a topic, with assessor markings and comments provided to improve knowledge which is hoped will lead to improvement of trainee performance (Black and Wiliam, 1998) and provide the maximum benefit to the trainee (Raymond and Usherwood, 2013).

Virtual Reality presents challenges in the ability to provide feedback of trainee experiences, as the trainee and assessors context and perspective may be different. The assessment conditions and methods have a direct impact on the ability to reliably provide feedback of trainee performances. This can be further exacerbated if the training or feedback is provided remotely. In face-to-face assessment the assessor must be present to observe the trainee in real-time during the simulation, simultaneously watching both the trainee operating within the Virtual Reality equipment, and the mirrored video output displayed on a screen (see Figure 1.3). Throughout the assessment, the assessor would likely be required to keep notes or monitor a checklist to determine performance outcomes, which would then be used as the basis for conducting assessment and providing the trainee with feedback.

Because the main advantage of virtual reality is the ability to conduct assessment and training anywhere, at anytime, it is important that feedback is provided in a timely manner (Wijewickrema et al., 2017). A streamlined approach that enables assessors to conduct authentic assessment and generate contextual feedback of the trainee performance is required. This work takes steps towards closing this gap by generating a framework suitable for assessors to assess room-scale virtual reality training simulations to conduct authentic assessment and generate contextual feedback.

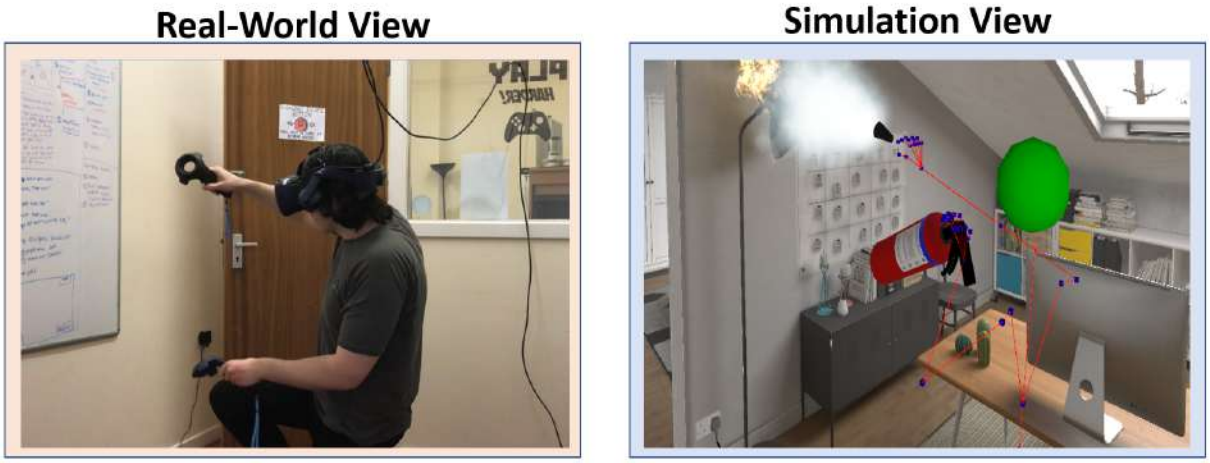


Figure 1.3: (Left) The view of the trainee in real-world using VR equipment. (Right) A third-person observation of the trainee in a virtual simulation environment.

1.2 Engaging with Swagelok Scotland

To close this gap, it was important to engage with industry experts who specialise in authentic assessment. As expressed by Adams et al. (2021), it is important to collaborate between disciplines, with emphasis on research being performed within both academia and industry to provide mutually beneficial teaching and learning. Swagelok Scotland specialise in education and training within engineering disciplines for providing qualifications to trainees to be certified to work on job-sites, such as is required to work within the oil and gas industry. Swagelok Scotland collaborated in this research to provide industrial insight into the requirements of assessors replicating classroom training and authentic assessment methodologies in VR.

The author collaborated in the development of a commercial VR training simulation with Swagelok Scotland external to his PhD, working on the development of an oil-rig training simulation. This simulation, which is featured extensively in this work, was designed and developed replicate the existing classroom methodologies for trainees performing authentic assessment. By transitioning the classroom authentic assessment methodologies, objectives and tasks into VR, trainees could operate within an immersive life-like environment that would be representative of the environment they would operate in when deployed for their job.

Transitioning the trainee experience into the VR simulation is a well established practice, with a large variety of simulations available within the commercial industry that replicate similar concepts (Wismer et al., 2021). However, upon completion of the VR simulation, it became evident that the their was no established concept or framework for transitioning the assessors experience into VR simulations. As authentic assessment is used during traditional classroom-based operations, it was identified that no solution was available to provide this comparable experience for the assessor, which relied on the ability to directly observe the trainee demonstrating their skills and knowledge within the VR simulation. Subsequently, this research correlated with the concerns of the company regarding assessors performing assessment for VR simulations, leading to their collaboration throughout this work for identifying the requirements and evaluating the framework, providing mutually beneficial findings that helped bridge the gap in knowledge and provided strategy for the company regarding future engagement with immersive technologies.

1.3 Definition of Key Terminology

In this work the following terminology will be used

- The term "*trainee*" refers to the someone who is training in a subject area to acquire knowledge or improve a performance or skill, such as a student or learner.
- The terms "*assessor*" refers to someone who has extensive knowledge and understanding of a relevant subject area (McCue, 2007). The term will be used to refer to both instructors and training assessors.
- The term "*assessment*" refers to the concept of measuring a trainee's progression and acquisition of knowledge and skills from trainee demonstrations (Deeley, 2018).
- The term "*feedback*" refers to the information supplied to a trainee after evaluation of a performance, providing constructive information that is hoped will lead to improved learning and re-enforcement of knowledge (Broquet and Dewan, 2016; Sadler, 1989).
- The term "*observation*" refers to the viewing of a event or scenario with visual and audio cues in either the real-world or in virtual reality (Jerald, 2016).
- The term "*deterministic*" refers to values or fields of data that are can be consistently reproduced when given the same input conditions and events.
- The term "*non-deterministic*" refers to values that are (or may) not be consistently reproduced when under the same input conditions and events.
- The term "*virtual reality*" refers to a computer generated digital environment which aims to replicate aspects of the real-world (Jerald, 2016).
- The term "*immersion*" refers to degree of stimuli a system and application provides to create a vivid illusion of the virtual environment (Slater and Wilbur, 1997).
- The term "*presence*" refers to sense of feeling located within a virtual space, tricking sensory information into believing the virtual space is real and letting users temporally forget the the real-world (Jerald, 2016).

- The term "*after-action*" refers to the process of reviewing tasks after the event has elapsed to reflect upon and discuss training performance connections. (Hanoun and Nahavandi, 2018).
- The term "*linear reconstruction*" refers to an experience that is reconstructed from data using an operation that requires a static and fixed approach to reconstructing data in order. This means actions must be reconstructed from the beginning of the recording matching the same format and order of inputs and actions captured. The data cannot be reproduced or reconstructed out of order, i.e. the reconstruction cannot be skipped ahead or reversed. The reconstruction must be restarted from the beginning if data needs to be re-examined.
- The term "*non-linear reconstruction*" refers to an experience that is reconstructed from data using an operation that is dynamic and allows a controllable approach to reconstructing data. This approach enables data to be reconstructed from any point of recording data, regardless of the order the data was acquired. Non-linear reconstructions can be skipped ahead, fast forwarded, or reversed without having to restart from the beginning.
- The term "*run-time*" refers to execution of an application that takes place live when a user (trainee) is present. This means the user (trainee) is active and interacting with the application or computer system.
- The term "*offline*" refers to the execution of an application that takes place when no user (trainee) is present. This means no user (trainee) is active, interacting with the application or computer system.
- The term "*recording instance*" refers to instances when variable data from a simulation has been saved.
- The term "*classroom-based*" refers to training, assessment or education performed within traditional real-world facilities.
- The term "*room-scale virtual reality*" refers to a virtual reality training simulation that operates using and/or requires a room-scale configuration.

1.4 Aims and Research Questions

The aim of this work is to address the following problems with regards to authentic assessment and contextual feedback of room-scale virtual reality training simulations and proposes a novel framework to address these problems:

1. There is a lack of empirical evidence and research focusing on assessors conducting authentic assessment and generating contextual feedback of room-scale virtual reality trainee performances.

2. There is a lack of empirical evidence for the tools and features required to be incorporated into a framework that enables assessors to conduct authentic assessment and generate contextual feedback of room-scale virtual reality trainee performances.

3. Recording run-time data of simulations can impact performance of the training experience, causing low frame-rate. If the simulation falls below the minimum frame rate, the trainee may experience motion sickness (Jerald, 2016)

1.4.1 Research Questions

This work takes steps towards closing the gap in knowledge for how a framework can be created that enables assessors to conduct authentic assessment and generate contextual feedback of room-scale virtual reality training performances. This core aim is split into seven distinct research questions:

RQ1 What is the state of the art for authentic assessment and contextual feedback of room-scale virtual reality training simulations?

RQ2 What evidence exists on usability and usefulness of identified approaches?

RQ3 What do assessors need for assessing and generating feedback for room-scale virtual reality training simulations?

RQ4 What tools are suitable to assess trainee performances in room-scale virtual reality training simulations?

RQ5 How can findings be compiled to create a framework for a generalised system for observation, assessment and feedback of room-scale virtual reality training simulations?

RQ6 Are the tools and features provided in the framework suitable for conducting authentic assessment and generating contextual feedback of room-scale virtual reality training simulations?

RQ7 How useful is the framework compared to state-of-the-art methods for conducting authentic assessment and generating contextual feedback of room-scale virtual reality training simulations?

These research questions are answered throughout the thesis, approaching each question in order of relevance to the research aim, as laid out in Section 1.5.

1.5 Proposed Approach

This research approaches the problems identified in Section 1.4 by designing and developing a framework that can capture virtual reality simulations and reproduce them for an assessor to conduct authentic assessment and generate contextual feedback to the trainee.

First, a literature review is conducted to discover the state of the art for methods, approaches and designs for assessing and providing feedback for virtual reality simulations. This literature review covers the methods for assessment and feedback, discussing each method usability and usefulness for conducting authentic assessment and generating contextual feedback, identifying whether there is a the gap in the literature. The literature review revealed that there is no framework that is suitable for authentic assessment and contextual feedback to be conducted for room-scale virtual reality training simulations.

Next, a study is conducted with industrial training assessors to identify current approaches to classroom-based assessment and training, and gather from their perspective, what they require to conduct authentic assessment and generate contextual feedback of room-scale virtual reality training simulations. These findings are then incorporated into tools and features requirements for the framework components. Finally, a framework is created that streamlines the process for assessors conducting authentic assessment and generating contextual feedback for room-scale VR training simulations, based on the findings from the literature and the requirements identified from the study with industrial training assessors.

The thesis concludes with an evaluation that analyses the suitability of the framework for conducting authentic assessment and generating contextual feedback of room-scale VR training simulations. This evaluation also compares the framework against two existing state-of-the-art methods discovered from the literature.

Before presenting the contribution of this work, the scope of this PhD work is discussed, justifying the approach taken.

Scope of Research

This work focuses on the perspective of assessors, whose role is to conduct assessment and generate feedback for trainee performances. To understand the assessors perspective, it is necessary to identify how authentic assessment and contextual feedback is conducted in classroom-based assessment and training, enabling the identification of requirements for the framework. These requirements form the basis for the tools provided within the framework, which is designed to provide assessors with the features necessary for conducting authentic assessment and generating contextual feedback of room-scale virtual training simulations.

The scope of this work focuses only on the perspective of assessors who operate the framework to conduct assessment and generate feedback. This research does not attempt to present findings from from the perspective of trainees. To enable the identification of components, tools, and features that are required as part of the framework, it was necessary to first focus this research on assessors. After the design, implementation and evaluation of the framework, it will be possible to expand the scope of work to examine the impact of authentic assessment and contextual feedback from the perspective of trainees.

With the evaluation of the framework presented in this work, other researchers and academics can build-upon this research, to evaluate other perspectives involved in the authentic assessment and contextual feedback process, such as the trainees.

1.6 The Framework

With virtual reality being viewed as a viable medium to complement or supplement training of real-world skill-sets (Backlund et al., 2007), a reliable framework that enables authentic assessment and contextual feedback of virtual reality training performances is required. As authentic assessment requires trainees to demonstrate a logical application of knowledge (Wiggins, 1990), assessors must be able to directly observe, monitor and assess trainee's virtual performances. Because automatic assessment is not yet capable of replacing human assessors (Wijewickrema et al., 2017), the framework must allow assessors to generate contextual and meaningful feedback. The framework proposed in this work is loosely based on the concept of Hanoun and Nahavandi's (2018) after-action methodology. Authentic assessment approaches trainees assessment based on their performance on the completion of intellectual tasks that resemble tasks they will encounter in real-world situations, rather than on abstract and indirect performance indicators (Wiggins, 1990). To perform this assessment trainees need to be observed completing the tasks. The after-action methodology (Hanoun and Nahavandi, 2018) allows assessors to observe trainees completing the task after the simulation has ended, enabling authentic assessment of the trainee performance.

To create a framework that is able to achieve the proposed objectives, three parameters must be considered: input, processing and output (Ramaprasad, 1983), or in the words of Hanoun and Nahavandi "1) Collect, 2) Diagnose and 3) Feedback" (Hanoun and Nahavandi, 2018, p. 1), with each parameter critical in the process of reaching the feedback stage (Ramaprasad, 1983). Figure 1.4 presents a high-level overview of how after-action review is conducted (Hanoun and Nahavandi, 2018), with the process divided into three categories: (a) the capturing of data, (b) the assessment itself, and (c) the generation of feedback for the trainee. The framework proposed in this work takes the generic stages in Figure 1.4 and develops them into specific components that are capable of enabling assessors to conduct authentic assessment and generate contextual feedback of room-scale VR training simulations.

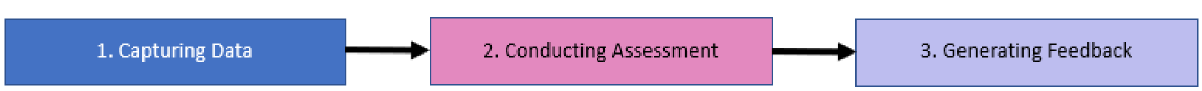


Figure 1.4: The framework has three very abstract stages for capturing data, conducting the assessment, and generating the feedback, as mentioned by Hanoun and Nahavandi (2018).

Expanding the stages in Figure 1.4 reveals the core outline necessary for the framework, while maintaining the same methodology for handling data.

Figure 1.5 shows an overview of the framework proposed in this work developed by expanding the stages in Figure 1.4 that achieves this research aims and addresses the research questions identified in Section 1.4. The focus of this framework is to incorporate an all-in-one approach to record, reconstruct, observe, assess, and provide feedback to trainee performances from virtual reality training simulations.

The stages of Figure 1.4 are overlaid onto the components of the framework in Figure 1.5. The role of the framework is not to replace the assessor, rather it is to make their job of assessing and providing feedback to the trainee easier, as suggested by Hanoun and Nahavandi (2018).

The framework consists of seven components, as illustrated in Figure 1.6:

1. Lightweight recording
2. Processed recording
3. Observation
4. Artificial intelligence
5. Live streaming
6. Assessment
7. Feedback

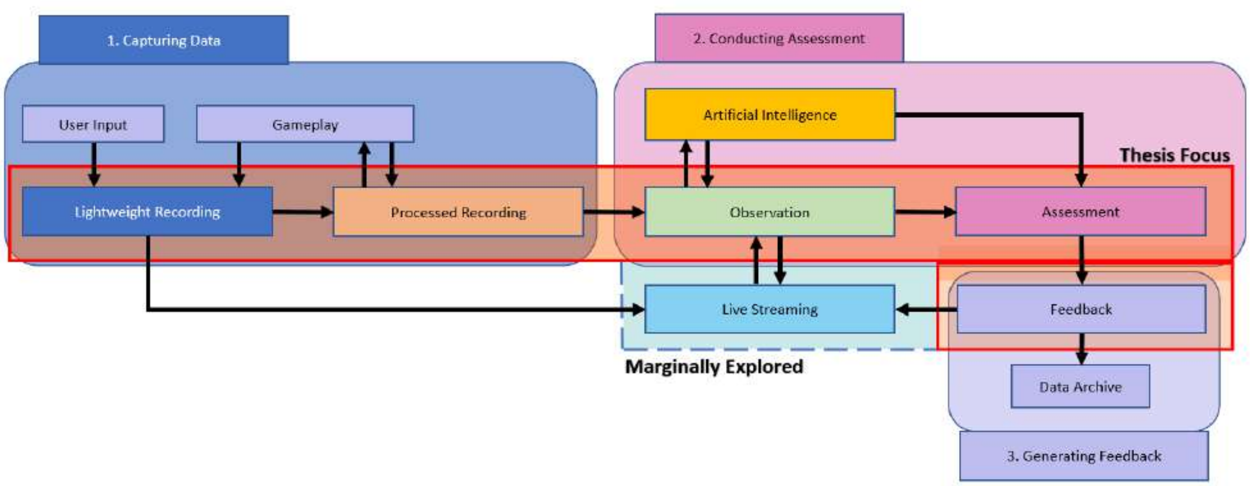


Figure 1.5: Overlay of parameters from (Figure 1.4) against the components of the framework, showing the connection between the components which are responsible for capturing data, conducting assessment and generating feedback.

Figure 1.6 shows the components of the framework without the overlaid stages from Figure 1.4.

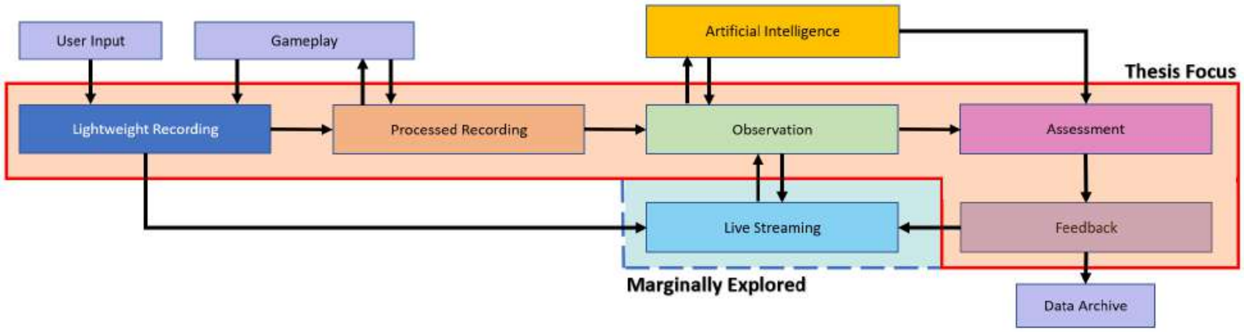


Figure 1.6: Overview of the framework necessary to conduct authentic assessment and generate contextual feedback of room-scale virtual reality training simulations. The components that are the focus of this work are enclosed within the orange box outline.

What separates this framework from existing methodologies of training performances, is that it adapts to room-scale VR. Because the VR trainee is not bound by conventional constraints for input such as keyboard and mouse, there is a wider of potential of actions possible within the simulation. This framework is generalised to capture all data from any type of simulation, including room-scale VR. By automatically capturing all data within a simulation, all assessors' concerns of the simulation failing to capture data, is removed (Hanoun and Nahavandi, 2018). The author's earlier work (Howie and Gilardi, 2019), demonstrated that it was possible for trainee performances to be consistently remotely recorded and reproduced using data from the simulation environment (see Section 2.7.7).

The framework uses a novel dual-recording methodology (Howie and Gilardi, 2021) to minimise disruption of simulation during run-time, splitting the recording operations to maintain a suitable frame-rate when the trainee is operating within the simulation. The *lightweight recording* and *processed recording* components are designed to capture all data from a trainee performance within a VR training simulation. This data is stored in a database, and enables reconstruction of the trainee performance for the *observation*, *assessment* and *feedback* components. Alternatively, the data can be used for deep-learning with *artificial intelligence*, or for *live-streaming* of trainee performances. However, these later two are outside the scope of this work and are not explored further.

To enable assessors to observe trainees, the observation component uses the data stored within the database to reconstruct the trainee performance. The reconstruction enables assessors to directly observe the trainee through reconstruction of the trainee actions and movement, in sync with simulation modifications. The assessment component goes beyond the visual and audible observation the trainee, embedding assisted assessment tools and features that allows for in-depth analysis and assessment of the trainee performance, enabling assessors to conduct authentic assessment. The final core component of the framework is the feedback component, which enables the assessor to generate contextual feedback.

1.6.1 Core Framework Components

This work focuses on the design, development and implementation of five fundamental components of the framework, highlighted in Figure 1.6, more specifically, lightweight recording, processed recording, observation, assessment and feedback. The dual-recording methodology, lightweight recording and processed recording components are designed and presented in Chapter 5, with the database, observation, assessment and feedback components designs discussed in Chapter 6. A technical implementation of the framework is detailed in Chapter 7.

Lightweight recording is the initial recording component of the framework presented in this work, which captures user (trainee) input and non-deterministic data, recording sufficient data that enables linear reconstruction of the trainee performance. Lightweight recording operates at run-time, simultaneously with the trainee operating the training simulation. The design of this component is detailed in Section 5.4, with technical implementation in Section 7.3.

Processed recording is the second and final recording component for the dual-recording methodology. Processed recording captures all remaining data within a simulation that was previously ignored during the initial lightweight recording. This remaining data is referred to as deterministic data, as it consists of all remaining elements of the simulation that, when given the same input, produce exactly the same data outcome. As such, processed recording component linearly reconstructs the trainee performance using the data captured by lightweight recording. During this reconstruction, all modifications to non-deterministic data are captured. The design of the processed recording component is discussed in Section 5.5 of this thesis, with technical implementation in Section 7.4.

The *observation component* presents the functionality of the framework that transforms the recorded data into a format which provides direct observation of the trainee. The observation component enables assessors to observe a reproduced trainee performance with complete control of their observation perspective and time. The observation component is detailed in Section 6.3, with a technical implementation in Section 7.5.

The *assessment component* extends upon the observation component and embeds assisted assessment tools and features into the assessors observation of the trainee. These tools and features include summative performance indicators, an augmented timeline and assisted visualisation tools, all of which are designed to aid assessors when conducting authentic assessment of trainee performances in VR. The assessment component is detailed in Section 6.5, with technical information discussed in Section 7.5.

The *feedback component* is represented as a combination of tools and features embedded into the framework that enable assessors to generate contextual feedback. The feedback component includes a video narration tool, evidence capture tool and drawing tool, which replicate classroom-based approaches for providing feedback. The feedback component of this framework is discussed in Section 6.6.

A publication of the author's earlier work (Howie and Gilardi, 2020) (see Figure 1.7), demonstrated the *live streaming* component. The live streaming component is envisioned as enabling multiple trainees or assessors to collaborate and communicate live in real-time. This can be used to enable live assessment and feedback, or allow assessors to support trainees with real-time communication. However, as this framework was intended for after-action assessment (Hanoun and Nahavandi, 2018), live streaming did not fit these requirements. As such, the live streaming component is considered as a supplemental component that can enhance the authentic assessment and contextual feedback capabilities of the framework, but it should not be viewed as necessary.

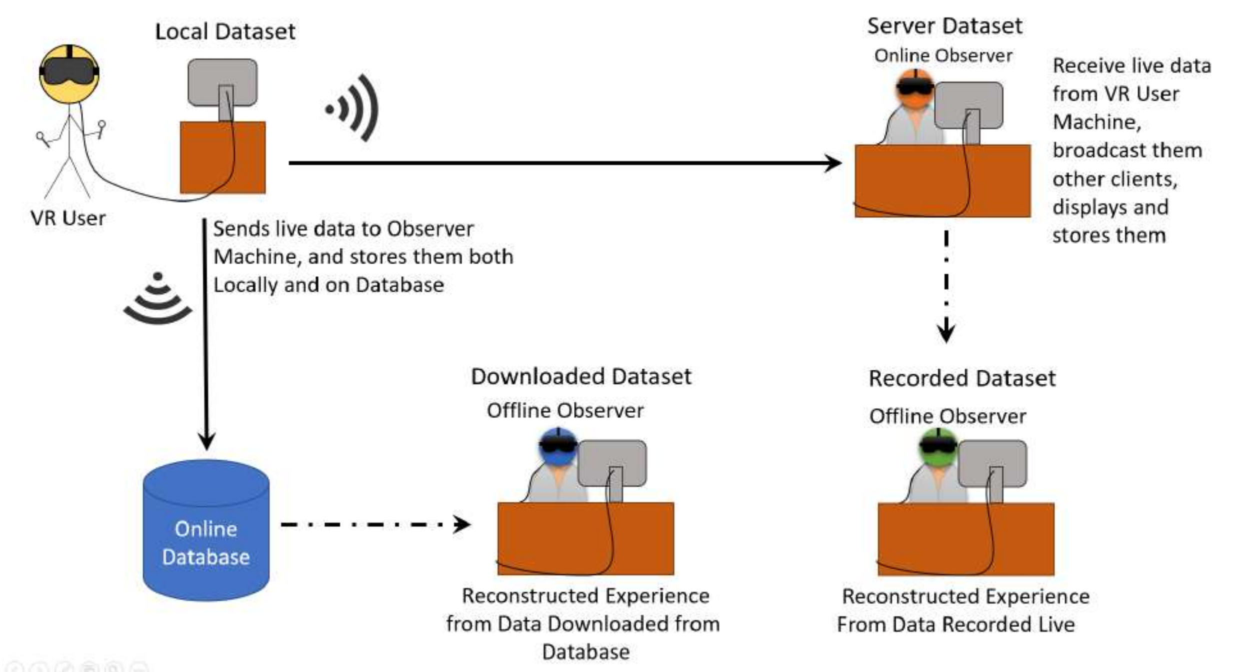


Figure 1.7: As presented in (Howie and Gilardi, 2020), the live streaming component enables instant observation of the trainee performances in real-time, along with having the transmitted data locally saved on the assessors computer or sent to a server for after-action review.

The *Artificial Intelligence* (AI) component is also not covered in this work. As the framework is designed to enable assessors to conduct authentic assessment and generate contextual feedback of room-scale VR training simulations, the artificial intelligence component is not required. As discussed by Wijewickrema et al. (2017), automated feedback is not yet sufficient to replace the expertise of an assessor. As such, the artificial intelligence component is seen a direction for future work. With advances in AI technology and deep-learning training models (Chilimbi et al., 2014), it is envisioned that AI will provide a useful assessment aid that could identify irregular patterns in trainee data. In this use-case it is envisioned that the AI component does not conduct the assessment, rather, it provides the assessor with deep insight into the trainee performance, further streamlining assessment and feedback.

The contribution to knowledge of this work focuses on the core components of: lightweight recording, processed recording, observation, assessment and feedback which are designed to provide assessors with the ability to conduct authentic assessment and generate contextual feedback of trainee performances in room-scale VR training simulations.

1.7 Contributions to Knowledge

The aim of this work is to design a framework that enables assessors to conduct authentic assessment and generate contextual feedback of trainee performances in room-scale VR simulations.

Contributions of this work are:

(a) A novel framework that enables assessors to conduct authentic assessment of VR room-scale training simulations.

(b) A novel dual-recording methodology, composed of two recording components. Lightweight recording operates at run-time, capturing trainee input and non-deterministic data. Processed recording follows, capturing all remaining deterministic data. As such, this approach has negligible impact on the run-time performance of the VR simulation.

(c) An all-in-one framework that embeds assisted assessment and feedback, tools and features, including: summaitive performance indicators, an augmented timeline, assisted visualisation of the trainee tools, video narration tool, evidence capture tool, and a drawing tool. These tools and features enable assessors to conduct in-depth assessment and analysis of trainee performances, and generate contextual and targeted feedback within a single framework environment.

This original work presents a requirement collection study with industrial assessors, followed by the design and development of a framework that enables authentic assessment of room-scale virtual reality training simulations. The originality of this work stems from the novelty of the framework and requirements necessary for assessors to perform authentic assessment in room-scale virtual reality.

1.8 Thesis Structure

In summary, this work describes and evaluates a framework that enables assessors to conduct authentic assessment and generate contextual feedback of trainee performances in room-scale VR. This thesis is organised as follows: In Chapter 2 a background literature is presented that discusses the background of virtual reality training assessment, including examining the usability and usefulness of existing methods of conducting assessment and feedback. Next, the methodology of the research and approach to answering the research questions is explained in Chapter 3. Following the methodology, a study with industrial training assessors is presented in Chapter 4, identifying the requirements for the framework components, tools, and features. The patent-pending dual-recording methodology is presented in Chapter 5, which leads into the design of the remaining framework components, tools, and features for observation, assessment and feedback in Chapter 6. Next, Chapter 7 presents a technical implementation of the framework using the Unity Game Engine. Chapter 8 presents the findings from the evaluation conducted with assessors. Finally, Chapter 9 concludes this work showing the contribution to knowledge and directions for future work.

1.8.1 Publications

Howie, S.R. and Gilardi, M., 2019, May. Virtual Observation of Virtual Reality Simulations. In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1-6).

Howie, S., Gilardi, M. Virtual Observations: a software tool for contextual observation and assessment of user's actions in virtual reality. Virtual Reality (2020). https://doi.org/10.1007/s10055-020-00463-5

Howie, S., Gilardi, M. Methods and Systems for Recording a User Experience. UKIPO Patent (Filed) GB2104444.1, March. 2021

Howie, S., Gilardi, M. Design and Development of Authentic Assessment Tools for Room-scale Virtual Reality Training. Virtual Reality (Submitted, expected 2021).

Chapter 2

Literature Review

The following chapter presents the state of the art for Virtual Reality (VR) assessment and feedback approaches. The methodology used to identify relevant literature is presented in Section 2.1. The learning cycle follows in Section 2.2, discussing Kolb's Experiential Learning Theory (ELT) and its incorporation within virtual training. Assessment and feedback and their role within learning is discussed in Section 2.3. The definition and implementation of authentic assessment is introduced in Section 2.4 evaluating its incorporation within VR training. VR immersion, presence and features of VR learning are discussed in Section 2.5.1. Existing implementations of VR training simulations are reviewed and analysed in Section 2.6, discussing its incorporation into training sectors and their recent transition from desktop to immersive VR. Methods of conducting assessment in VR training simulations for after-action review are discussed in Section 2.7. Methods of generating and providing feedback for trainees within VR training simulations follows in Section 2.8. Finally, Section 2.9 summarises the findings from the literature highlighting the gap in knowledge for authentic assessment and contextual feedback of room-scale VR training simulations.

2.1 Review Methodology

An archival methodology was been used to identify relevant literature from the following databases: Association for Computing Machinery (ACM) Digital Library, Science Direct, Education Resources Information Center (ERIC), Scopus, Web of Science Service for UK Education, Springer, BioMed Central, Institute of Electrical and Electronics Engineers (IEEE), Sage Journals and PubMed Centeral (PMC). This search was augmented by a recursive search of the identified literature, inspecting references to further expand the discovery of relevant research. Papers relevant to the aims of this work were analysed based on initial impressions of the title and abstract, followed by conclusion, and the introduction, finishing with the references. If the content of these sections were deemed irrelevant, the paper was discarded. The search of literature included the following keywords and terms:

("virtual reality" OR "game based learning" OR "e-learning")
AND assessment
("virtual reality" OR "game based learning" OR "e-learning")
AND feedback
("virtual reality" OR "game based learning" OR "e-learning")
AND training
("virtual reality" OR "game based learning" OR "e-learning")
AND observation

Relevant papers were identified by title and then abstract. Based on the abstract selection, the relevant papers were included in the literature. Selection of identified papers was determined by their discussion, involvement and relevancy of the assessment and feedback process for learning, along with papers relevant for training, assessment and teaching within Virtual Reality or game-based learning. The literature search was not restricted by decade or time-frame. Only three papers were identified as directly relevant for the development of this work, namely (Hanoun and Nahavandi, 2018), (Stone, Snell, and Cooke, 2016) and (Raij and Lok, 2008), because this low number of directly related papers a narrative approach was chosen to present the results of the literature review. In the narrative literature a broad view of assessment and feedback is given, discussing papers useful for the development of a framework for authentic assessment in room-scale VR even though they are not directly related to the development of assessment VR tools or frameworks. This narrative literature review informs, critique and analyse the state-of-the-art methods for assessment and feedback in VR.

Table 2.1 presents the search results for each term for assessment, feedback, training and observation. Duplicate papers found within search terms were removed.

Table 2.1: Search results from each database using the search terms. Numbers present total search results, with the numbers enclosed between brackets indicating the number of papers that were identified as relevant.

Total Search Results				
Database	assessment	feedback	training	observation
ACM	21,454 (6)	21,011 (9)	21,380 (4)	21,369 (3)
Science Direct	41,092 (5)	39,644 (6)	42,546 (3)	36,771 (0)
ERIC	1,101 (4)	585 (2)	1,281 (1)	341 (0)
Scopus	47,136 (1)	46,853 (1)	46,915 (1)	46,895 (2)
WSSUKE	50,967 (0)	49,536 (3)	53,025 (0)	48,293 (1)
Springer	35,098 (4)	38,238 (8)	52,222 (11)	27,723 (4)
IEEE	42,637 (2)	42,305 (0)	43,296 (6)	41,805 (5)
Sage Journals	11,412 (2)	5,203 (0)	11,499 (2)	10,786 (2)
PMC	41,897(5)	30,267 (5)	41,733 (4)	21,648 (1)

In the following chapter, the literature review has been structured to establish a comprehensive, critical and objective theoretical framework for the research, setting the foundations for future chapters of this work. The first element in this narrative structure is to provide a background understanding for authentic assessment and discuss how it relates to virtual forms of training.

2.2 The Learning Cycle

As this work analyses training within VR simulations, it is important to understand how training is established within learning. VR simulations provide an opportunity for experiential learning, in-which trainees operate a simulated experience with forms of engagement that require involvement and acts of "doing". The best-known method of experiential learning is Kolb (1984) Experiential Learning Theory (ELT). Kolb (1984) considers the acquisition of knowledge as "the process whereby knowledge is created through the transformation of experience" (Kolb, 1984, p. 38).

Kolb (1984) developed ELT to provide an understanding for how we learn, grow and develop from experiences. The model demonstrates how learning is acquired from experiences, and provides a structure for how learning can be crafted to repeatedly refine learning and acquisition of knowledge (Kolb, 1984). Kolb's (1984) ELT contains four stages, concrete experience, reflective observation, abstract conceptualisation and active experimentation, which must be followed in sequence (Figure 2.1), but can be initially entered from any stage (Kolb, 1984). To achieve a complete learning experience, all four stages of ELT must be experienced by the trainee. These four stages are as follows (Kolb, 1984):

1. Concrete Experience: Focuses on a task in-which the subject has active involvement and experiences, such as practical exercises that involves a form of engagement for "doing".

2. Reflective Observation: Detaches and steps away from involvement and the act of "doing", and focuses on reflecting upon the experience to review and contemplate engagement. An example being, self-reflection or discussions of the experience with others.

3. Abstract Conceptualisation: Is the process of evaluating and making sense of the experience to interpret the relationship between what has just been experienced, and how it fits into what is already known, which can be drawn from developed knowledge or theory from textbooks. This can be achieved by presenting ideas or writing a paper that conceptualises the learning outcomes from the experience.

4. Active Experimentation: Determines how learned knowledge can be applied in practice, taking the understandings of knowledge and applying them to predicate outcomes of actions. This final stage then repeats, leading aback to concrete experience. It is important that the knowledge is applied within context to a task, otherwise, their is a risk the knowledge acquired will be quickly forgotten.

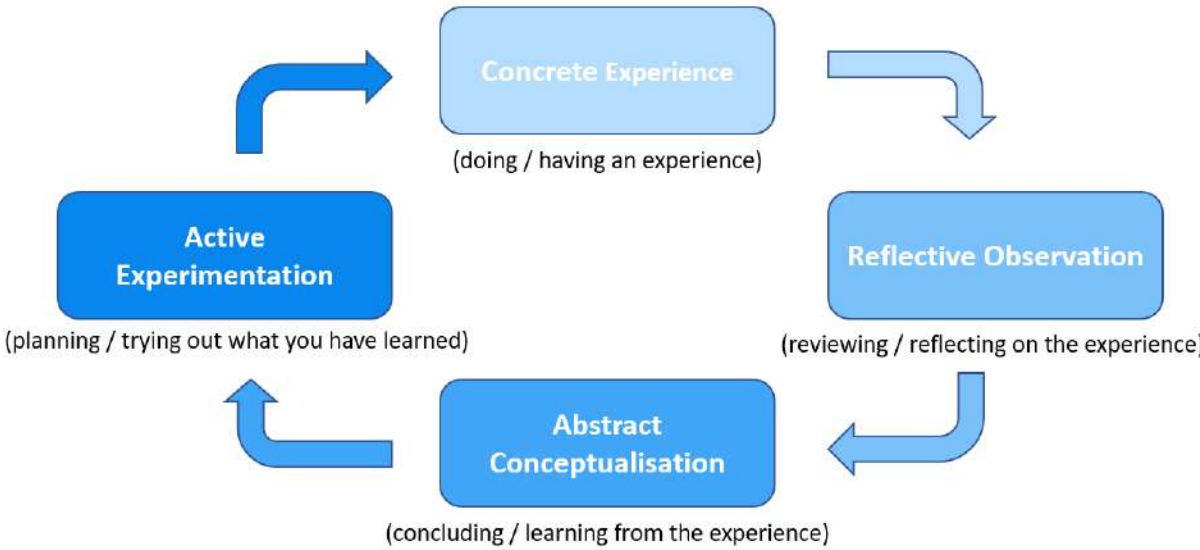


Figure 2.1: The four stages of Kolb's (1984) Experiential Learning Theory.

Kolb's (1984) ELT remains relevant within virtual forms of learning, with many studies incorporating the model into the learning process of simulations or virtual studies (Kolb, 1984; Chen, Toh, and Ismail, 2005; Krishnan et al., 2019). As such, the four stages of ELT present a model that can be replicated within VR training, forming the background understanding for how trainees use VR experiences to acquire knowledge and improve their skill through reputation of involvement, reflection, conceptualisation and experimentation, otherwise referred to as experiential learning (Kolb, 1984).

2.3 Assessment and Feedback

Before reviewing how assessment and feedback can be conducted in virtual reality training, it is important to clarify the elements of assessment and feedback that are incorporated into real-world learning scenarios.

2.3.1 Assessment

Assessment focuses on the progression of knowledge and skills through learning of a subject matter (Wang, 2018; Deeley, 2018). Assessment is vital to provide trainees the opportunity to demonstrate their skills, so that an expert can rectify misunderstandings which will hopefully lead to improvement in future demonstrations (Deeley, 2018), with effective assessments generating further discoveries and learning for a trainee performance (Abraham and Singaram, 2019). Shute, Ke, and Wang (2017) makes clear that assessment is an important element within the learning process, as it provides information regarding trainee understanding and identification of performance, which can be used to establish the trainee's current knowledge for a given subject area. However, assessment is not restricted to traditional classroom-based work, and is incorporated in all forms of physical and virtual learning. One of the most prominent sectors has been e-learning, which has demonstrated usefulness for providing assessment and feedback opportunities (Deeley, 2018) in virtual life-like scenarios. These virtual scenarios conduct assessment by allowing a trainee to demonstrate their knowledge via application, with the assessment outcomes determined by the completion of the learning outcomes and objectives (Konakondla, Fong, and Schirmer, 2017; Carruth, 2017). However, there are many forms of assessment, and not all apply the same methodology for reviewing the performance or work of a trainee, with an emphasis on the perspective of assessment and form of assessment. The three core methodologies for conducting assessment are discussed in this Section.

Self-assessment as the name suggests, is a self-evaluation of an individuals work, requiring the individual to take proactive role in the assessment their own performance, providing themselves with feedback (Nicol and Macfarlane-Dick, 2006). However, for self-assessment to be effective the learner must clearly understand what good performance is, and identify the validity of their own performance in relation their knowledge of good performances (Sadler, 1989; Nicol and Macfarlane-Dick, 2006). Furthermore, the individual must understand the background knowledge and skills required to improve their own performance and reach perfect results (Sadler, 1989). For self-assessment to be effective, learners need opportunities to create knowledge and awareness of their subject matter to determine and position their own work with respect to the wider community (Nicol and Macfarlane-Dick, 2006). In virtual forms of learning, self-assessment is hard to achieve as learning could be conducted remotely or independently of a skilled expert, making it difficult to create opportunities for the learner to develop self-assessment abilities required to reliably assess their own work. As demonstrated by Andersen et al. (2019), self-assessment did not provide satisfactory results for learning within virtual formats, requiring the support from assessors to continue a trainee's skill development.

Summative assessment is a summarising of performance indicators. Summative assessment is often utilised in e-learning forms of assessment as it is easy to generate feedback since the outcomes focus on predefined scores and grades (Fardinpour, Reiners, and Wood, 2018). For example, summative assessments are frequently designed as multi-choice questionnaires, or closed-answer questions that have predefined limitations of outcomes which the participant must select from, simplifying the assessment and marking (Fardinpour, Reiners, and Wood, 2018). However, due to the limited scope of responses, summative assessments only provide an overall impression of a participants knowledge, lacking sufficient in-depth information required to provide directive feedback (Fardinpour, Reiners, and Wood, 2018). Summative assessments focus on generating a summary of participant performance and achievements that are quick and easy to identify (Sadler, 1989). However, because summative assessments are designed to generate an overall score, the process causes examiners to simplify the results, implying that assessments ignore rich formative data of individual responses or actions (Fardinpour, Reiners, and Wood, 2018).

Summative assessment is frequently found in generalised areas of virtual training and e-learning, mostly due to their easy to digest nature that can quickly generate a score as a learning outcome, providing a visible grade of overall knowledge for certification (Goderstad et al., 2019; Nicol and Macfarlane-Dick, 2006). However, the overall outcome score is not an accurate or reliable form of assessment for critical learning. In the case of virtual training, the objective is to provide immersive training to prepare trainees for real-world scenarios. While summative assessment provides digestible pass or failure decisions, their meaning has little substance of usefulness for revealing rich-insight into trainee problems or concerns, lacking the context to understand how or why the trainee passed or failed a task.

Formative assesment is an in-depth assessment of participant skill for each task or objective, determining the quality of a trainees work (Sadler, 1989). Unlike summative assessment which simplifies the learning outcomes, formative feedback requires an expert with knowledge of the assessment domain to accurately determine the suitability and validity of trainees work (Fardinpour, Reiners, and Wood, 2018). In comparison with summative assessment, formative assessment is difficult to evaluate outcomes using metrics, and instead relies upon the quality, application and concepts used to perform a task (Sadler, 1989). Furthermore, formative assessment is often utilised as a method for further learning, i.e. the contents of the assessment produced enables further opportunities for learning (Deeley, 2018). In turn, formative feedback is considered as a method of providing information regarding the variability of success and failures of individual performances conducted a participant (Sadler, 1989).

Because authentic assessment requires focus on the trainees quality and validity of work, formative assessment is best suited for this role. With a formative assessment methodology, the applied work conducted by a trainee can be assessed step-by-step by a skilled expert, allowing for direct feedback for correcting individual mistakes, therein-by creating learning by doing, followed by corrective learning from feedback (Deeley, 2018). Furthermore, with formative assessment, Shute (2009) state that assessment should support, rather than undermine the learning process of an individual. With formative assessment, feedback can be focused on providing directed formative feedback that targets the problems faced by the trainee, rather than providing generic support (Shute, 2009). For assessment to be complete it is important to supply appropriate feedback.

2.3.2 Feedback

The purpose of feedback is to provide a trainee constructive information that is hoped will lead to improved learning and re-enforcement of knowledge (Broquet and Dewan, 2016; Sadler, 1989). Feedback is a critical element of the learning process, providing the trainee with useful and usable information that re-enforces their performance outcome with directions for further learning. However, while assessors are usually responsible for providing feedback to trainees, it falls upon the trainee to utilise the feedback to enable an improvement of their skill-set or knowledge to occur

(Sadler, 1989).

Feedback supports the learning and education of individuals by increasing motivation in the learning process, with assessment used to compare and judge the current trainee performance to an expected performance level for feedback to be generated (Sadler, 1989; Nicol and Macfarlane-Dick, 2006). Feedback can be categorised as intrinsic or extrinsic. Intrinsic feedback is self-administered, which consists of self-assessment of an individual performance, with aims to reflect and learn from their experience (Wang, 2018). Extrinsic feedback is external feedback provided by a trainer or expert in the subject area (El Boghdady and Alijani, 2017; Wang, 2018).

While feedback is considered one of the most important tools available to educators (Broquet and Dewan, 2016; Wijewickrema et al., 2017), it is best provided as a supplemental tool that expands upon the knowledge acquired from training exercises, allowing reflection (Wijewickrema et al., 2017; Billings, 2012), comparison of performance(s) (Fardinpour, Reiners, and Wood, 2018), correction of mistakes and misunderstandings (Wijewickrema et al., 2017; Billings, 2012; Davis, 2005). It is also important that feedback is effective in correcting relevant actions performed by a trainee, providing targeted formative comments that praise strengths and highlight weaknesses of the individual under assessment (Fardinpour, Reiners, and Wood, 2018).

Broquet and Dewan (2016) note that effective feedback includes information that can be both positive and negative, with referencing to a performance objective to indicate when and where feedback is applied. If feedback is generalised and not targeted to the individual, the feedback lacks relevance to the tasks and objectives of an individual trainee performance. Generalised feedback does not provide adequate guidance for trainees to rectify their problems or learn from their mistakes (Broquet and Dewan, 2016). In the same vein, feedback and assessment of information should not only consider immediate results, but be utilised long-term to determine changes in performance (Ellaway and Masters, 2008) and identify if mistakes are frequently repeated to indicate a gap in knowledge. As a consequence assessment information should be stored in a centralised database that adds reliance to avoid loss of data (Ellaway and Masters, 2008) and limits human errors in manual storage of data.

It is crucial that feedback is provided to trainees in a timely manner, since delayed feedback risks becoming redundant (Jonsson, 2013), with overdue feedback preventing the trainee from learning and improving prior to subsequent assessments (Jonsson, 2013). Furthermore, it is important that feedback is unique and targeted to the flaws of the individual, as generic feedback risks being unhelpful for the trainee to learn from (Jonsson, 2013).

2.4 Authentic Assessment

Authentic assessment is the theory of examining performance based on the completion of intellectual tasks, rather than on abstract and indirect performance indicators (Wiggins, 1990). The ability of VR training simulations to reproduce real-world situations lend itself to authentic assessment approaches. Newmann and Archbald (1992) describes authentic assessment as a tool to provide high-order thinking and problem solving capabilities for the individual and society. VR training simulations apply authentic assessment to demonstrate application of real-world knowledge during meaningful training and assessment exercises (Wood et al., 2013; Lee, 2017). Conducting authentic assessment involves the use of authentic tasks that are relevant to the context that determine trainee learning outcomes (Gulikers, Bastiaens, and Kirschner, 2004), this is often because jobs focus on authenticity of performance and competencies (Gulikers, Bastiaens, and Kirschner, 2006). Unlike assessment of authentic achievement, where emphasis is placed on the learning outcomes, authentic assessment examines the authenticity for the manner of the assessment (Joy Cumming and Maxwell, 1999). However, performance demonstrations between contexts are not indicative of each other, as the domain knowledge could be contextual to a relevant learning scenario (Joy Cumming and Maxwell, 1999).

In higher education, there is a detachment from academic forms of assessment and practical applied assessment, with academic problems often artificially contrived and detached from real problems (Boud, 1995). Consequently, recommendations have been made to move to a competency and authentic assessment process, which is naturalistic, authentic and well-constructed (Boud, 1995). As the name suggests, authentic assessment attempts to evaluate the validity of performance that simulates real-world viability, and be meaningful for the logical application of knowledge (Wiggins, 1990).

There are several key aspects that separate authentic assessment from traditional forms of standardised testing. For authentic assessment, trainees must be effective at applying and adapting acquired knowledge to a task. The variety of performance tasks form challenging activities that requirement engagement by the trainee to solve. With this engagement, authentic assessments provides a vehicle for the trainee to demonstrate and justify their knowledge for overcoming challenges. As such, authentic assessment provides a valid and reliable method for measuring performance and knowledge of the trainee. In summary, authentic assessment tasks are authentic to the challenges and roles of real-life problems and situations, preparing trainees for a transfer of knowledge that is adaptable to real-world situations (Wiggins, 1990). In contrast, traditional forms of assessment are limited in scope for questions and answers, and are used to regurgitate answers learned within contexts. Traditional forms of assessment focus on the answer supplied by the trainee, often ignoring the reason and decision making that lead to the answer, with testing focused on drilling arbitrary information in an abstract format from real-life, adapted from Wiggins (1990).

In essence, authentic assessment should be used to support a trainee in a way that is forward thinking for real-world application of knowledge (Wiggins, 1990), unlike traditional outcomes which use scores that "have no obvious meaning or usefulness [, which] undermine teachers' ability to improve instruction and trainees' ability to improve their performance" (Wiggins, 1990, p. 1). Authentic assessment represents an opportunity for trainee's to demonstrate their skills and knowledge in a manner that is observable and provides evidence of competency within real-world challenges (Wiggins, 1990).

In authentic competency-based assessment, tasks must reflect the competency needs being assessed, which reflect authentic real-world tasks, and the application of knowledge is required to complete the assessment task (Gulikers, Bastiaens, and Kirschner, 2004; Gielen, Dochy, and Dierick, 2003; Joy Cumming and Maxwell, 1999). With these conditions, authentic competency-based assessments have higher validity for measuring the competency of the trainee compared to traditional assessment approaches. Gulikers, Bastiaens, and Kirschner (2004) define authentic assessment as an assessment of competencies that would apply in a professional setting. Ashford-Rowe, Herrington, and Brown (2014) (Ashford-Rowe, Herrington, and Brown, 2014) agrees, declaring that the challenges of assessment should reflect the authenticity of real-world objectives and situations. Authentic learning provides a context for constructive and situated problem solving and knowledge construction (Wood et al., 2013).

Collins (2002) and Joy Cumming and Maxwell (1999) argue that hands-on experience can only be acquired using laboratory or real-world experiences, the former being clinical, with the later being confrontational. These forms of experience cannot be learned through lecturers, and require a hands-on approach to facing the role, with the need to bridge the gap from academic knowledge to real-world experiences (Collins, 2002). The role of virtual reality provides the avenue for direct activity experiences, supplementing traditional training to provide an in-place learning experience. However, it is crucial that assessment and feedback of performances is comparable to that gained from real-world confrontations, both from within the simulation and assessor feedback.

Virtual training simulations can represent a high fidelity experience with authentic assessments, but safe training conditions may not replicate the dangerous real-world scenarios which stress trainee competency in real situations (Gulikers, Bastiaens, and Kirschner, 2004). Trainee's can experience cognitive overload if placed within a real or complex environment, negatively impacting the learning outcomes (Sweller, Van Merrienboer, and Paas, 1998). As such, virtual simulations can provide the needed abstraction from real-world conditions, while still fulfilling the necessary conditions for training, tailoring learning to trainee abilities and needs (Gulikers, Bastiaens, and Kirschner, 2004). In authentic assessment, the performance, procedures and outcomes are all considered part of the assessment, with the simulation used as a vehicle to demonstrate trainee capabilities (Messick, 1994).

While virtual reality provides a route for authentic assessment, Gulikers, Bastiaens, and Kirschner (2004) and Joy Cumming and Maxwell (1999) expresses doubts as to whether the conditions of the virtual context are comparable to real-world conditions, given the context is safe for the trainee. It is preferred that authentic assessment is conducted in real environments, such as work-placements, however, this is not often possible, and in most cases it is the exception rather than the rule (Wood et al., 2013). While simulations can be viewed as framed artificial imitations of the real-world which avoids unforeseen complications or issues (Joy Cumming and Maxwell, 1999), the advancements in technology and simulations are constantly pushing the boundaries and capabilities of immersive simulations, as witnessed by capabilities and depth of aviation training simulations (Lee, 2017).

Authentic learning is the process in which knowledge will be utilised in real-world, engaging in the same logical application of knowledge that would be expected in a real-world scenario (Wood et al., 2013). Wood et al. (2013) discuss that while the gamification elements of simulations appear on the surface to be inconsistent with authentic assessment , the tasks, setting and assessment remain authentic, with gamification elements available to support the trainee learning. The inclusion of gamification elements enable the learner to practice free of constraints, practicing and repeating process while remaining faithful to the authentic learning objectives (Wood et al., 2013) It is however important that continuous formative feedback is supplied to trainees when using authentic assessment (Murphy et al., 2017). Feedback should follow, with repeated evaluation and reflection, reviewing trainee performance and assisting to build upon their knowledge for future application of knowledge (Murphy et al., 2017).

Ashford-Rowe, Herrington, and Brown (2014) state that the process of providing, receiving and discussing feedback is vital should be included in any authentic assessment activity, as a indicator of performance for future improvement. In this process, the learner should reflect on their experience, followed by discussion of their feedback to improve learning (Ashford-Rowe, Herrington, and Brown, 2014). Simulations are often deployed to learn and evaluate the use of real-world skills and competencies, allowing for improved development of higher level skills compared to traditional approaches (Vos, 2015). Vos (2015) found that the vast majority of authentic assessment use-cases in education literature for simulations utilised formative assessment and feedback provided by an assessor. This supplemented the learning provided by the simulation itself, providing better development opportunities for trainees (Vos, 2015).

Qandil, Darweesh, and Al-Ghananeem (2021) believes that real-world examinations can transition online using simulations and authentic assessment. However, while assessment and feedback are at the core of measuring authentic assessment (Ashford-Rowe, Herrington, and Brown, 2014; Vos, 2015; Murphy et al., 2017) remote online assessment poses a challenge, with the need for suitable learning management systems and assessment platforms that provide the necessary tools and features to support learning (Qandil, Darweesh, and Al-Ghananeem, 2021), further highlighting the current need for the framework presented in this work.

In conclusion, it is important to understand how virtual reality training simulations function, with a focus on the training procedures most commonly adopted. However, first it is important to clarify the hardware required to enable room-scale virtual reality training.

2.5 Virtual Reality Hardware

To enable room-scale virtual reality experiences, the VR equipment requires a form of room-scale tracking capabilities (Jerald, 2016). This can be achieved using wall mounted sensors for Vive equipment, or through an optical tracking system that estimates the equipment position using the cameras located on the headset, an example being the Oculus Quest lineup of equipment. The use of tracked controllers provides new capabilities for interaction and immersion within virtual environments (Jerald, 2016). To replicate training methodologies applied within classroom-based scenarios, it is important that trainees are able to naturally apply their skills and knowledge. As such, most modern state-of-the-art VR equipment provides tracked controllers which offer room-scale tracking capabilities when paired with suitable VR equipment (Jerald, 2016). When determining an interaction device, it is important that the device and interaction method are equally considered LaViola et al., 2017. The device must be equally calibrated to the task, and the task must be equally calibrated to the device, picking the device that is based suited to a interaction technique, and then modifying that interaction technique to best fit the devices properties and characteristics LaViola et al., 2017. The accessibility of 3D printing technology also provides new avenues for user created equipment (LaViola et al., 2017). DIY devices may be developed to create stronger connections to the physical prop and virtual object, creating additional equipment required for a training exercise, which enables concepts like subsitutional reality (Simeone, Velloso, and Gellersen, 2015), which use props to map physical interactions into the VR simulation environment Cordeil et al., 2017.

One of the main advantages for using wall mounted sensors is the ability to scale the tracking devices, with capabilities for extending the equipment to include additional sensors, such as the Vive trackers, which allow for full body tracking. This additional tracking equipment is utilised for the evaluation in Chapter 8, providing the ability for observation of an estimated trainee skeleton structure.

While these concepts are shared within the field of XR technologies, this work focuses on the concepts that apply within the area of VR for room-scale immersive environments.

2.5.1 Virtual Reality - Definitions

There is a lack of accepted consensus for the term *immersion* (Agreval et al., 2019). Within this work the definition of immersion given by Jerald (2016), i.e. the degree of stimuli a computing system provides to create vivid illusion of the virtual environment (Jerald, 2016), is adopted. Immersion can be related to a trainee's psychological state and connection with the properties of technology (Agreval et al., 2019), which encapsulates a system that provides an experience (Agreval et al., 2019).

Immersion requires a match between the trainee's proprioceptive feedback of body movements and collation of information generated and displayed in the world (Slater, Usoh, and Steed, 1995).

The two accepted perspectives of immersion reference an individuals psychological state, or the objective properties of the technology system used (Agreval et al., 2019). Brown and Cairns (2004) highlight that immersion is not possible if a simulation suffers from usability through control problems, relying on seamless control interaction to achieve total immersion. When total immersion is achieved, the trainees are detached from the reality of the real-world, being completely engrossed within the simulation (Brown and Cairns, 2004). The content of the virtual environment plays a vital role in the experience of immersion (Agreval et al., 2019). It is anticipated that trainees will require several minutes to achieve a state of immersion (Agreval et al., 2019), however it is difficult to quantify and measure immersion, as trainee mood, preferences and emotional states can impact their response to questionnaires (Agreval et al., 2019). When trainees are fully immersed in an VR simulation, they experience *presence* (Jerald, 2016).

Presence is a psychological state that is difficult to define, and is best understood when experienced (Jerald, 2016). Presence alters the perceptions and psychological awareness of the real-world with virtual replacements (Jerald, 2016), described as an accumulation of sensory information that provides a sensation of ‘being there’ (Jerald, 2016; Slater et al., 2009). This sensation is achieved by the unknowing replacement of sensory feedback through equivalent virtual stimuli (Slater et al., 2009), creating temporally unawareness of the real-world, accepting the objectives, events and characters presented within a virtual environment (Jerald, 2016), described as the feeling of being lost within the experience (Agreval et al., 2019).

Sensory cues can enhance the feelings of presence within virtual environments (Cooper et al., 2018), with trainees who experienced a high feeling of presence, performing best within VR simulations (Cooper et al., 2018). However, presence often constrained by the immersion provided of the VR equipment or simulation (Jerald, 2016), with the feeling of presence being affected by the amount of control one has over their interactions within the virtual environment (Schloerb, 1995). If presence is lost, the illusion of the virtual environment and simulation objectives are realised by the trainee, breaking the immersion of the VR experience (Jerald, 2016).

The literature discussed shows the importance of *immersion* and *presence*, which can be experienced simultaneously (Agreval et al., 2019), within virtual simulation environments, highlighting that these physiological impressions can drastically alter the trainee’s attitude towards the simulation and virtual environment. Without the feeling of presence, trainee’s may not perform tasks authentically, knowing the virtual environment is not real, as discussed by Gulikers, Bastiaens, and Kirschner (2004) in Section 2.4. As such, simulations should attempt to authentically reproduce virtual environments in a format that provide natural, authentic and well-constructed scenarios that enable authentic assessment (Boud, 1995).

2.5.2 Virtual Reality Learning

The use of VR as a form of learning provides the opportunity to exposure trainees to new scenarios, challenges and experiences, without any risk of danger (Jensen and Konradsen, 2018; Hamilton et al., 2021). This Section will discuss a brief overview of the advantages and uses of VR for training and learning.

Chavez and Bayona's (2018) systematic literature review of 30 papers for VR in learning between 1999 and 2017, found that 97% of author's identified interaction capabilities as necessary for successful implementation of learning within VR characteristics, with a further 73% discussing the importance of tools suitable for providing immersion (see Section 2.5.1). Recent publications topics have emphasised the role of assessment and evaluation within the principles of VR in education (Chavez and Bayona, 2018). VR learning has been shown to have positive effects within the learning process, by leading to improved learning outcomes, providing experiences closer to reality, increasing curiosity and interest compared to traditional academic learning, and the acquisition or improvement of skills (Chavez and Bayona, 2018).

Hamilton et al. (2021) systematic literature review identified 29 recent publications within the scope of learning in VR from 2013, revealing that in around half of cognitive studies, VR learning demonstrated a positive learning benefit when compared to non-immersive practices. The least effective studies were found to have utilised multiple-choice questionnaires, which stifled trainee demonstration abilities for measuring learning outcomes (Hamilton et al., 2021). In procedural tasks, which allowed trainee demonstration of knowledge via application of procedures to complete objectives, two studies (Yoganathan et al., 2018; Sankaranarayanan et al., 2018) demonstrated a transfer of knowledge, revealing that VR learning improved results for real-world application (Hamilton et al., 2021).

While most studies have shown VR learning producing positive results (Jensen and Konradsen, 2018; Chavez and Bayona, 2018; Hamilton et al., 2021), there are still substantial barriers that need to be overcome, such as motion sickness, lack of appropriate software, technical and hardware limitations (Jensen and Konradsen, 2018). Furthermore, Jensen and Konradsen (2018) argues that while VR can be utilised as a method of training, it does not add value, as it is more expensive. Jensen and Konradsen (2018) also fear that the virtual environment will distract from the learning process. However, these arguments do not account for the difficulties and costs associated with industrial or collaborative real-world training, as discussed in Section 1.1, or the requirement of equipment, which VR learning avoids (Hamilton et al., 2021). Although it is possible task irrelevant stimuli can initially distract from the learning process in the short-term, familiarity and positive exposures to the technology and simulation will rectify these distractions (Jerald, 2016).

The main criticism of VR learning is the assessment methodology often adopted, with Hamilton et al. (2021) arguing that learning outcomes should go beyond analytical outcomes and Chavez and Bayona (2018) demonstrating the characteristic advantages of VR as a learning medium that shows improvement over traditional academic measurements. The literature discussed shows that the use of authentic assessment for measuring trainee competency in VR simulations is justified, enabling trainees to demonstrate knowledge logically, rather than relying on abstract academic indicators (Wiggins, 1990) (see Section 2.4), which most existing VR learning simulations fail to achieve (Hamilton et al., 2021). However, this is often due to the inability to assess in-depth trainee performances, with Hamilton et al. (2021) stating that VR learning simulations require evaluation and assessment tools that can provide a complete overview of the learning achieved within VR simulations. These discussions further highlights the gap in research of a framework suitable for conducting authentic assessment within virtual reality training simulations.

In summary, the common sentiment from the literature indicate that VR education positively effects the learning process and outcomes (Chavez and Bayona, 2018; Hamilton et al., 2021), with the positives out weighting any potential negatives.

2.6 Virtual Reality Training Simulations

Training is essential for the acquirement of knowledge for many professions (Kandemir, Soner, and Celik, 2018). However, in real-world it is not always possible to conduct training due to a multitude of reasons and methods for different sectors and industries. For instances, concerns regarding safety are often a key issue, since some training may endanger participants, putting trainee health and safety at risk (Fardinpour, Reiners, and Wood, 2018; Latham et al., 2019; Carruth, 2017).

Alternatively, business concerns are another common reason behind lack of specific training (Fardinpour, Reiners, and Wood, 2018; Bucarelli, Zhang, and Wang, 2018), mostly emphasised by the cost and time required to conduct training not being deemed worthwhile by management compared to the risk imposed. For example, natural disasters are frequent, but it is not economically and practically viable to prepare real-world training for all scenarios (Fardinpour, Reiners, and Wood, 2018; Farra et al., 2019). Rather, training is used to emphasise focus on probable disasters (Fardinpour, Reiners, and Wood, 2018; Farra et al., 2019).

Because real-world training is not always viable (Latham et al., 2019; Carruth, 2017), Virtual Reality training has been used to subsidise or supplement real-world training due to its reduced cost (Farra et al., 2019; Carruth, 2017), accessibility (Farra et al., 2019; Hsu et al., 2013), safety (Roy, Bakr, and George, 2017), ability to be assessed (Carruth, 2017; Andreatta et al., 2010), on-demand training (Andreatta et al., 2010; Ragazzoni et al., 2015) and collaboration potential (Hsu et al., 2013; Ragazzoni et al., 2015). These methods combine to making VR training a reusable and consistent assessment framework (Fardinpour, Reiners, and Wood, 2018; Ellaway and Masters, 2008). In some instances, VR simulations have been found to equal or better traditional laboratory training (Gunn et al., 2018) and shown to be effective (Farra et al., 2019).

Virtual simulators can be designed to create authentic looking virtual worlds (Ellaway and Masters, 2008; Carruth, 2017). Kandemir, Soner, and Celik (2018) states that simulated environments are one of the most effective tools for assessment, providing real-world experiences (Kandemir, Soner, and Celik, 2018; Carruth, 2017). Immersive VR training provides a method for learning by doing through engagement, with an action hands-on approach preferred to passive viewing, listening or observing (Abidi et al., 2019; Fardinpour, Reiners, and Wood, 2018; Latham et al., 2019). Unlike traditional forms of assessment, virtual reality training provides a method of assessment that enables measurement of trainee competency for application of knowledge to complete a task (Fardinpour, Reiners, and Wood, 2018). Immersive experience can also enhance the learning experience due to the visual realism achievable, creating a safe and immersive training experience (Latham et al., 2019). Virtual training simulations allow people to experience immersive life-like scenarios with the objective of transferring the virtual learning experience to real-world knowledge safely (Kandemir, Soner, and Celik, 2018). For example, flight simulators utilising realistic props for cockpits and interfaces to create realistic interactions using virtual experiences (Fardinpour, Reiners, and Wood, 2018).

There are many variations of VR based training. Desktop VR has been frequently used for years in serious-games, with notable use in platforms like Second Life (Bredl et al., 2015; Danforth et al., 2009; Inman, Wright, and Hartman, 2010). Other forms of virtual training is also evident in fields such as maritime simulation training (Charissis et al., 2008a; Stone, Snell, and Cooke, 2016), driver training (Bozkir, Geisler, and Kasneci, 2019), chemistry education (Abuhammad et al., 2021), and health/medical training (Moore et al., 2020; Raij and Lok, 2008; Charissis et al., 2008b). While, not all of these papers incorporate immersive VR, the trend of recent literature suggests that training is transitioning away from desktop VR and into immersive room-scale VR experiences (Moore et al., 2020; Berndt et al., 2018; Shaw et al., 2019; Bayerl et al., 2019), as room-scale VR offers significantly higher immersion and presence (Buttussi and Chittaro, 2017).

In this move to immersive VR for room-scale experiences, VR simulations may require additional tracking capabilities to validate the authenticity of assessment and applied knowledge. For example, a comparison by Bernard et al. (2018) compared a VR simulation with real-world mock-up, finding significant differences in trainee postures between the mediums. The VR group was found to have utilised an unsuitable posture when completing certain objectives within the simulation (Bernard et al., 2018). These postures would not be possible for conducting the task in the real-world due to physical constraints missing from the virtual scenario (Bernard et al., 2018). Findings from (Bernard et al., 2018) indicate that it may be necessary to track the posture of trainee's within VR training simulations to determine if their stance is acceptable.

Another avenue of using VR training is the ability to reach remote audiences and enable them to experience life-like simulations for the purpose of educating, training and learning. In this approach, charities have started utilising immersive VR technology to educate people of the dangers that exist during natural disasters (Red Cross, 2018). Rather than focusing on training for staff and employees, training simulations can be deployed to train, teach and educate civilians, raising the awareness and preparedness of people in at-risk locations. Other gamified VR training scenarios exist that utilise the same premise, but deployed on consumer storefronts that allow for remote at home training and education (LinderoEdutainment, 2017).

VR simulations can also be used to collaborate as teams in multi-user scenarios for group work or global co-operation. Klomp, Spitalnick, and Reissman (2011) highlighted the importance of the VR simulation replicating the conditions and situations of real-world training scenarios. As such, Klomp, Spitalnick, and Reissman (2011) allowed for two participants to simultaneously co-operate within a single training session, to allow trainee's to gain valuable experience co-operating, reducing anxiety and pressure involved when deployed in a difficult situation. Trainees evaluated the simulation experience, demonstrating that they had been able to apply learned skills from the VR training to real-world situation (Klomp, Spitalnick, and Reissman, 2011).

Klomp, Spitalnick, and Reissman (2011) highlighted the importance of training workers within their expected job capacity, deploying trainees in teams if expected during a real-scenario (Klomp, Spitalnick, and Reissman, 2011). When creating their simulation scenario, they noted that most virtual environments do not support collaboration, and lack viable methods intuition for co-operating with their team members, which validates similar findings from other studies (Bliss, Tidwell, and Guest, 1997). These papers expose the need for multi-user training when replicating real-life scenarios, highlighting the need for training communication skills, collaborative team problem solving and co-ordination of tasks (Bertram, Moskaliuk, and Cress, 2015).

The use of multi-user VR training introduces further challenges for assessment and feedback, as group projects require monitoring of both the group as a whole, and each individual (Alruwais, Wills, and Wald, 2018). Monitoring trainee's in these scenarios requires knowledge of communication between collaborators, and detailed information of work conducted by each trainee (Alruwais, Wills, and Wald, 2018; Ellaway and Masters, 2008). Without suitable assessment and monitoring tools, it is difficult to reliably judge group work, it is therefore difficult to provide suitable and useful feedback within multi-user VR training simulations (Alruwais, Wills, and Wald, 2018).

Concerns regarding assessment and feedback is significantly important since VR is is growing as a viable medium for training. For effective assessment and feedback to be conducted, the wide variety of fields that operate VR training needs to be considered.

In summary, immersive room-scale VR training is becoming mainstream, and is being used to educate and assess the trainee competency, skills and knowledge, through demonstration(s) of authentic assessment. This training should be conducted in a format that enables in-depth assessment and feedback of their performance. However, many VR training simulations have only limited assessment capabilities (Moore et al., 2020; Eubanks et al., 2016; Buttussi and Chittaro, 2017), which are often restricted to analytical data, which only provides pass or fail outcomes. With most simulations not providing capabilities for direct observation or assessment assessors, it is difficult to measure the authenticity of trainee performances, as the application of knowledge can not be reviewed or validated. Furthermore, for simulations that do capture trainee performance for direct observation, it often requires manual recording of the performance using video captures of the desktop or immersive screen (Bayerl et al., 2019).

Solutions for conducting assessment and feedback is discussed in the following sections.

2.7 Assessment of Virtual Reality Training Simulations

Unlike real-world training, virtual forms of training alters the context of observing, evaluating and assessing trainees performance. Virtual forms of training are useful for learning since they provide an objective platform for assessment that can be tracked and built-upon to improve learning without risks to the trainee (Konakondla, Fong, and Schirmer, 2017; Carruth, 2017). With the advancement of virtual technology, assessment capabilities for virtual forms of training are undergoing changes, with advances in technology driving the ability to capture authentic, complex and skill-based assessments (O'Leary et al., 2018).

However, during the design and development of virtual reality training simulations, the assessment and learning outcomes are often secondary considerations. Therefore, there is a concern that the learning may be based on "faulty paradigms" that do not determine if the true learning outcome is achieved (Raymond and Usherwood, 2013). Consequently, to measure the true learning outcomes of simulations, assessors should be incorporated into the design and development of the virtual training simulations to define the expected outcomes (Raymond and Usherwood, 2013). Assessment within the virtual training simulations can then be designed to ensure the tasks and objectives defined are true learning outcomes which measure the trainee performance (Raymond and Usherwood, 2013). However, it is important that feedback follows assessment by incorporates elements of debriefing and feedback that reference the trainee performance (Raymond and Usherwood, 2013).

Assessment and feedback in simulations are sometimes considered a time-consuming burden by assessors (Raymond and Usherwood, 2013; Pietquin, Hastie, et al., 2013; Gould et al., 2001), however, it is an invaluable source for maximising learning from the training simulations. As a consequence, it is vital that assessors are continuously consulted during the design and development of training experiences to ensure the learning outcomes to be conducted by trainee within a simulation are of good standard (Fardinpour, Reiners, and Wood, 2018). By incorporating assessors into the development process of the simulation, there should be no confusion regarding the learning outcomes and objectives. With true learning outcomes represented within the simulation, assessor who assess the work have a viable baseline to judge and assess the trainee performance (Sadler, 1989)

Assessment is important in virtual training simulations as it provides evidence of the trainees current knowledge (Shute, Ke, and Wang, 2017; Johnson and Priest, 2014; Hattie and Clarke, 2018), which can be used as a platform for providing feedback. Pedram et al. (2020) investigated the impact of explanatory feedback on trainee learning outcomes using an immersive VR training simulator for mine rescue, explaining trainees strengths and weaknesses. Pedram et al. (2020) findings suggest that the assessors had a direct impact on the learning outcomes of the trainees, concluding that explanatory feedback has a positive outcome on learning process, stating that it should be considered an essential component for any training program (Pedram et al., 2020). Pedram et al. (2020) findings are validated by the literature, with Martins and Kellermanns (2004) and Wan and Fang (2006) presenting similar arguments. Pedram et al. (2020) also note that that trainees can learn and benefit from self-reflection of past training experiences and useful feedback provided by skilled assessors. Johnson and Priest (2014) argue that for novices, explanatory feedback is more effective than corrective, as explanatory feedback explains the reasons behind their success or failure, whereas corrective feedback only informs of a success or failure (Pedram et al., 2020)

The communication between trainee's and assessors allows for opportunities to correct mistakes and misunderstandings, and further improve learning in trainee knowledge (Pedram et al., 2020). Furthermore, Pedram et al.'s (2020) findings are in line with those discussed previously in Section 2.3, suggesting that traditional and virtual methods of assessment and feedback are closely related.

Automatic assessment of complex skills or rich data is still not viable and may take years still to match or surpass the input from an assessor (O'Leary et al., 2018; Beller, 2013; Bennett, 2015). More recently, Armas, Tori, and Netto (2020) continues the discussion, reinforcing that automatic assessment is not yet suitable for VR simulators and remains a challenge still to be overcome. Therefore, most instances of automatic or instant forms of assessment rely on analytical calculations that are easier to configure (O'Leary et al., 2018). Whereas, assessment of virtual reality training simulations will require human intervention due to the nature of authentic assessment. Consequently, the assessment of a training performance should be assisted by the computer system to intelligently aid human involvement during after-action review process (Johnson and Gonzalez, 2008; Hanoun and Nahavandi, 2018).

Armas, Tori, and Netto (2020) summarised that simulators lacked automatic assessment capabilities and prevented a complete evaluation of participant actions during performance. While some simulators offered the ability to evaluate limited features of a trainee performance, they did not cover in-depth assessment capabilities of a trainee full interactive experience (Armas, Tori, and Netto, 2020). This highlights the importance of involving assessors in the assessment and feedback stages, enabling them to provide trainees with direct feedback using their expertise knowledge as an additional learning process of the training experience (Sadler, 1989; Broquet and Dewan, 2016; Ellaway and Masters, 2008). However, because this research focuses on after-action assessment, it is important to understand the methodology and approach for assessment before evaluating existing assessment approaches.

2.7.1 After-Action Assessment

Assessment as described in this work refers to an after-action review and analysis of a trainee performance. As discussed in Section 1.6, the process of after-action assessment includes three stages: (1) the collection of data, (2) the diagnoses, and (3) the feedback provided to a learner (Hanoun and Nahavandi, 2018). Subsequently, Hanoun and Nahavandi (2018) considers video recording, multiple camera view playback, bookmaking of events with timestamps and virtual info overlay to rendered scenes as the most effective features for collecting and assessing trainee performances.

Hanoun and Nahavandi (2018) discuss recommendations for incorporating after-action review assessments in virtual training simulations:

1. The assessment process should be decoupled from the training environment to enable easy re-use.
2. Data collection must function without any errors or be impacted by human bias.
3. Multiple sources of performance data is required to guarantee correctness.
4. Measurement of performances must be based on the domain knowledge of the trainee.
5. Comparison of performance must be based on expert performance.
6. Comparisons should determine the quality of performance between the trainee and expert.
7. The assessment and feedback must provide valid explanations that provides the trainee with the ability to understand what their strengths and weaknesses are.
8. Assessment and feedback should be conducted immediately after the training to keep the experience fresh in the mind of the trainee.

However, while it is recommended assessment and feedback is conducted immediately after training, it is unlikely that this recommendation is viable for most use-cases, since a primary advantage of virtual reality is the ability to provide remote and distant forms of learning and training (Xie et al., 2018). Although the time delay between a trainee conducting an assessment and the assessment and feedback being conducted by an assessor should be minimised, it is improbable that instant assessment and feedback is possible with human forms of assessment. As such, it could be determined that a hybrid assessment and feedback approach is best suited, providing some from of instant automated feedback to the trainee, which can then be supplemented by assessor generated feedback (Hanoun and Nahavandi, 2018).

In most after-action tools for conducting assessment, the methods share the same core capabilities, with the focus on collecting data that can be replayed to observe and analyse a training exercise or to generate trainee reports (Johnson and Gonzalez, 2008). These tools often include features such as book marking (timestamps), time manipulation and offer an objective perspective of the training exercise (Johnson and Gonzalez, 2008)

The next section identifies existing methods of assessment from within the literature. Because there are many different approaches and existing methods for conducting assessment in virtual reality training simulations, such as; analytical data, motion and input, stealth-based assessment, action based, animation recording, environment recording, video assessment and game engine implementations. Each method has been categorised to a unique section that is relevant for its methods of capturing and presenting trainee data. The remainder of this section will review, compare and discuss each method and their suitability for room-scale virtual reality training simulations, starting with analytical data.

2.7.2 Analytical Data

Analytical data uses simulator data values to determine the outcome of a simulation, which can vary in scope and complexity. In the most primitive form of analytical feedback, the data collected can be the presentation of analytical performance values, such as completion time action identifiers (Carruth, 2017), which are insufficient for effective evaluation (Pan et al., 2015). To overcome the limitations of analytical data, Carruth (2017) detailed a technique that captures input and movement logs from the trainees and additional actors within a simulation, creating a recording of the actions (combination of movement and input) and interactions conducted. This recording can be used to replay and analyse the responses and actions taken by a trainee. However, the assessment tool is designed to operate independently and provides automatic assessment and feedback using pass or fail conditions from pre-determined values. These outcomes are determined by the validness and accuracy of the learner application of knowledge acquired from the log data (Carruth, 2017). Carruth (2017) utilised the replaying functionality to validate the values used for generating the pass or fail outcomes, rather than rely on assessors, for fear human assessors would be frustrated linearly watching the entire trainee performance. As such, replaying would require the ability to for assessors to focus on the important areas of the performance for validation. These time concerns partially drive the ambition for automatic assessment and for the incorporation of more direct and accessible methods conducting assessment.

To drive analytical assessment of simulations, there are several metric calculation values, such as: time completion, number of errors, motion data, objectives completed, along with additional variable data unique to a training simulation (Lahanas, Georgiou, and Loukas, 2016; Konakondla, Fong, and Schirmer, 2017; Jensen et al., 2017; Bric et al., 2016; Moore et al., 2020). These analytical values may be captured automatically by the software experience (Jensen et al., 2017; Bric et al., 2016), or require an assessor to record data and score parameters (Goderstad et al., 2019). The measurement of time as a metric is widely utilised in judging trainee performance, as it is easy to compare against a target time value (Carruth, 2017). However, time is not a reliable method of acquiring insight into the knowledge captured by assessment, since external methods within simulation can interfere with the participation. Likewise, presenting metrics as performance target values may lead to trainees operating the virtual system to achieve goals in faster times, rather than operating in normal safe and correct procedures, resulting in poor transfer and application of knowledge for real-world uses. A potential solution is to appropriately weight the timing metrics with all other important critical metrics (Bric et al., 2016), but correct balancing of all these metric values is difficult to achieve, due to the scope and complexity of virtual training simulations.

Chaudy (2015) developed "EngAGe", a learning assessment engine for assessment of educational (serious) games. EngAGe (Chaudy, 2015) offers developers and educators the potential to adapt games learning outcomes for trainees, and provides a platform that presents statistically and visually the analytical metric data captured. Using a web platform, EngAGe provides educators a method to access and modify simulation data to alter the scoring of objectives during training and learning, providing educators with direct access to the trainee metric data (Chaudy, 2015). This provides educators and assessors insight into the scoring and engagement of trainees, however, unlike Carruth's (2017) assessment tool, EngAGe does not provide the ability for direct visual observation of the trainee. In room-scale or complicated virtual training simulations, the metric data is insufficient for effective evaluation (Pan et al., 2015) as it does not provide a true picture of the trainee experience. However, this does not reduce the prevalence analytical metric forms of assessment in virtual training simulations.

In summary, the main advantage of analytical data is the ability to provide a consistent platform for assessment (Jensen et al., 2017), which can be used by automated assessment systems to provide instant and automatic outcomes for the trainee performance (Jensen et al., 2017). Analytical methods are useful for creating generalised outcomes of trainee performances, notably the ability to identify novice and expert participants (Jensen et al., 2017) based on the score differences. However, while analytical data provides a scored metric outcome which appears easy to setup and quick to understand, the data lacks the depth necessary for authentic assessment. Therefore, analytical data is best suited to summative assessment of trainee performances, in which the role of assessment is to generate pass or failure outcomes, rather than formative insight into the individual actions of a trainee during their performance.

2.7.3 Motion and Input Tracking

Motion and input tracking is used to generate data from input devices used within a trainee performance, acting as a an extension of analytical data. However, the approach extends beyond internal simulation logic and data, incorporating the ability for tracking input device values. A prominent example is the inclusion of heart rate monitors and other sensors (Salgado et al., 2018; Raij and Lok, 2008) during a simulation. The inclusion of external data can sync with the simulation data to provide richer information of the simulation learning outcomes of a trainee (Salgado et al., 2018).

There are various types of motion and input tracking methods available, but the most prominent implementation is the work by, MacCormick et al. (2019), who created a full body evaluation system for Virtual Reality called FRVRIT, which is only compatible with HTC Vive (HTC Vive, 2019) equipment. This system runs in the background during third-party games and is designed to be a generic motion and input replay system for any VR game (MacCormick et al., 2019). The data from FRVRIT is used to generate a visualisation of the VR operators movement against a timeline of the recorded motion, with options available to alter the VR input devices being monitored (HMD, tracker or controller). An observer can use any of the seven predefined camera perspectives to view the trainee motion in from the VR equipment, allowing direct observation and visualisation of the movement of the trainee (MacCormick et al., 2019) However, as acknowledged by the author's (MacCormick et al., 2019), FRVRIT is only useful for visualisation of movement, and lacks any context to sync the movement of the trainee within a simulation experience (MacCormick et al., 2019). For future work, the author's indicate they intend to track the VR operators vision during gameplay of third-party games for future iterations of the method (MacCormick et al., 2019).

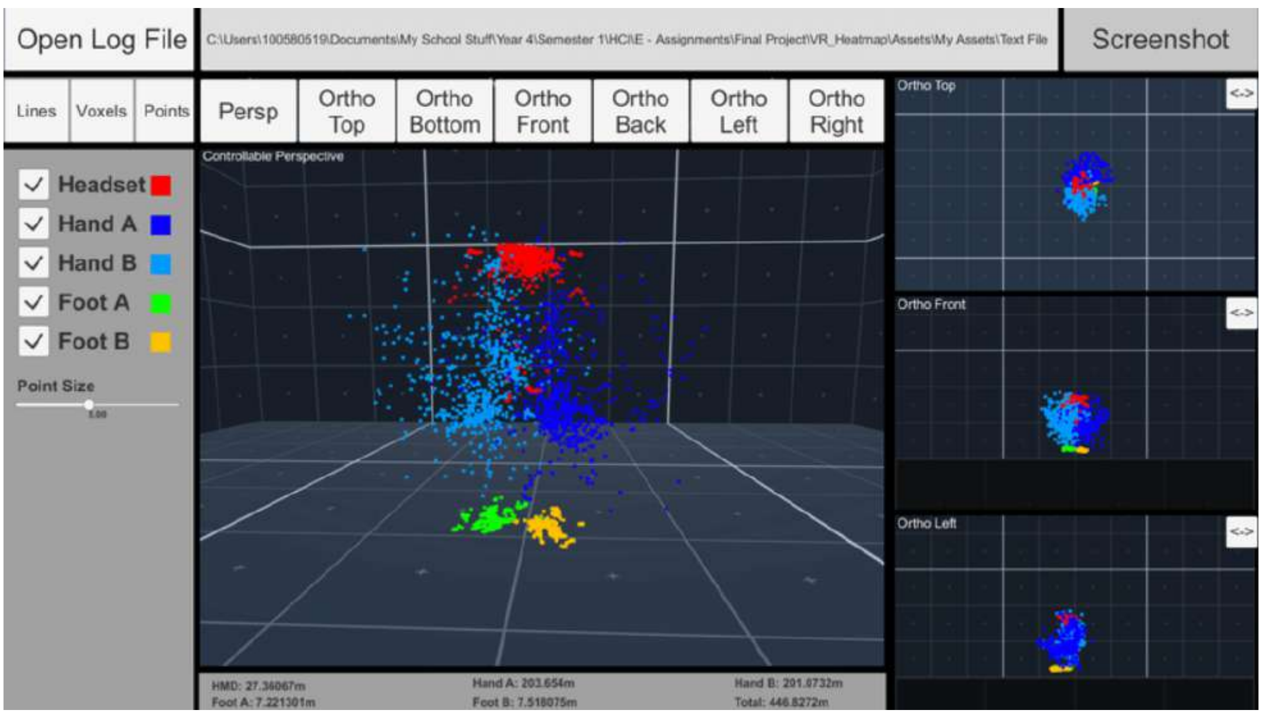


Figure 2.2: MacCormick et al.'s (2019) FRVRIT tool displaying the motions of a participant playing the video game Beat Saber (Image published with author's permission).

Motion and input tracking is an interesting approach to collecting and analysing data in virtual training simulations. However, because the approach shares too many similarities with analytical data, it lacks sufficient context for the assessment process. While the input and motion tracking of a VR trainee provides insight into trainee input and motion, the approach lacks knowledge and context regarding what the trainee is actually doing within the simulation experience (MacCormick et al., 2019). If the approach offered a method of direct observation, the combined metric information could be used to enable in-depth analysis of the trainee performance (Salgado et al., 2018). Future iterations of FRVRIT are suggested to incorporate some form of context (MacCormick et al., 2019), however, because the data collection is not embedded into the simulation, syncing the internal logic of the simulation with FRVRIT is unlikely to be feasible. Because FRVRIT captures the external VR hardware data from input, the tool does not have access to any simulation logic or variable information from the simulation. Therefore, it is difficult to perceive how a method for providing simulation context to the data could be implemented, while remaining generic enough to suit all virtual simulations.

2.7.4 Stealth-Assessment

Stealth assessment is a technique of assessment which is conveyed as both assessment and learning, embedding the assessment process into the learning objectives of the simulation (Shute, 2011; Shute et al., 2021). Stealth assessment is hidden into the functionality of a game experience which continuously acquires session data during active engagement with a trainee (Shute, 2011). Being hidden, stealth assessment overcomes one of the concerns with evasive assessment approaches which introduce pressure and anxiety from being continuously observed (Shute, 2009; Shute, Ke, and Wang, 2017). While participating in game sessions, trainees are tasked with completing objectives. To complete objectives successfully, trainees must conduct a set of actions using acquired knowledge to overcome gameplay challenges, employing critical thinking and problem solving skills (Shute, 2011). The actions conducted by trainees provide a rich set of data that creates an activity of their competency to overcome the challenge. However, the data from stealth assessment is designed to identify findings from trainee data that reduces the work of assessment, rather than to providing a method of enabling observation of the trainee performance By gathering data throughout a session, stealth assessment attempts to determine the ways and methods a trainee reacts and interacts during gameplay (Shute, 2011; Shute et al., 2021). For example, stealth assessment captures the steps and actions a trainee conducted to overcome a problem within a simulation (Shute, 2011). Gameplay actions inform the competency model based on what is being assessed (Shute et al., 2021), these are classified with performance analytic (good/bad) which provide evidence for an estimated competency rating of a trainee (low, medium or high) (Shute, 2011; Shute, Ke, and Wang, 2017; Shute et al., 2021).

Using stealth assessment, automatic assessment of data can be achieved in a limited capacity. The process of understanding participant responses to problems, creates a possibility for a suitable response to be automatically and instantly be generated and provided to the trainee as feedback. Shute (2011) demonstrate automatic assessment and feedback using a evidence-centered design, competency model and Bayesian networks approach (Pearl, 2014) in which a trainee response can be compared against an expert response(s) to provide feedback on misconceptions or mistakes. Shute's (2011) system can be used to measure trainee performance indicators to provide scores, such as the trainees novelty or efficiency to complete an objective (Shute, 2011). These attributes can be acquired from other learner data or supplied by assessors, with novelty for example, determined by the frequency of trainee approaches to an individual problem (Shute, 2011).

An advantage to using stealth assessment with competency models is the ability to streamline expert engagement with learner data, simplifying the grading task and providing a foundation for directed formative feedback from an assessor (Shute, 2008; Shute, Ke, and Wang, 2017). Findings from stealth assessment can be used to adapt the training experience to the skill-level of the learner, by providing additional hints or modifying the difficulty of the session (Shute, Ke, and Wang, 2017). However, stealth assessment shares similarities to analytical data and motion and input tracking, as discussed in Section 2.7.2 and 2.7.3. Consequently, the focus of stealth assessment is to assist the assessment process for the assessor, which is the ultimate goal of human focused assessment methods (Johnson and Gonzalez, 2008). However, because stealth assessment is focused on identifying relationships within the data, it does not consider any method of reproducing the trainee performance to enable direct observation.

2.7.5 Action Based

Action based assessment determines learning outcomes by examining the actions of the trainee against a criteria of expected actions, allowing trainees to demonstrate their skills and knowledge for overcoming objectives (Fardinpour and Dreher, 2012). In most simulations and learning outcomes, if the participant does not complete an objective, it is considered to not have achieved learning outcomes(Fardinpour, Reiners, and Wood, 2018). However, in some cases this can be counter-productive to the learning process, especially if an error is only conducted at the very end of the simulation, while all previous steps were successfully completed (Fardinpour, Reiners, and Wood, 2018). Action based assessment monitors and assesses smaller sequence of events that lead to sub-objectives in a simulation (Fardinpour, Reiners, and Wood, 2018). Therefore, action based assessment is able to determine where and how a trainee failed during a simulation by identifying irregularities in the actions completed.

Fardinpour, Reiners, and Wood's (2018) action based assessment operates by comparing participant actions against the predetermined actions of assessors. During simulations, activities conducted by the participant are captured as raw data against time of the simulation (Fardinpour, Reiners, and Wood, 2018), creating timed action related information. Fardinpour, Reiners, and Wood's (2018) implementation of action-based assessment does not evaluate trainee performance automatically, as it requires an assessor to create a formula to weight and measure the assessment, along with providing explanatory feedback (Fardinpour and Dreher, 2012). The assessment outcome is determined by the compliance of rules, weight of attributes, timing of attributes, achievement of goal and sequence of actions related to expert predefined actions (Fardinpour, Reiners, and Wood, 2018). Actions can be deemed as optional or essential, allowing small divergences in the expected actions conducted by participants (Fardinpour, Reiners, and Wood, 2018). However, if an essential action is missed, the outcome would still be considered an overall failure (Fardinpour, Reiners, and Wood, 2018).

Reiners, Wood, and Dron (2014) implemented a narrative based assessment which follows a sequence of events that can be compared and assessed to determine outcomes, sharing many similarities in approach to Fardinpour, Reiners, and Wood's (2018) work. While Reiners, Wood, and Dron's (2014) work does not explicitly define itself as action-based, its implementation shares many similarities with learning objectives of Fardinpour, Reiners, and Wood's (2018) work, with both implementations allowing divergence in actions conducted as long as the outcome is achieved (Reiners, Wood, and Dron, 2014). In Reiners, Wood, and Dron (2014) divergence is allowed to provide the trainee curiosity with the environment, allowing trainees to be immersed and engaged during tasks and objectives, as long as the engagement is focused on progression towards the learning objective (Reiners, Wood, and Dron, 2014). The narrative structure focuses the engagement of the trainee, allowing flexibility in tasks with modelled predefined outcomes configured to determine if actions conducted by the learner are within the scope of the objective (Reiners, Wood, and Dron, 2014). However, because action-based assessment compares the expected actions to predefined expert actions (Fardinpour, Reiners, and Wood, 2018; Reiners, Wood, and Dron, 2014), action based assessment systems can only determine the reliability of actions following the expected flow, as such, unexpected actions or nuances conducted by the trainee will go unnoticed.

Both Fardinpour, Reiners, and Wood (2018) and Reiners, Wood, and Dron (2014) action based assessment systems share the same process, define anticipated actions with a skilled expert, categorise critical and optional objectives, and then generate a formula which determines the learning outcomes and assessment ability. For their purpose, the assessment is viable, however there are several issues that impact its viability for room-scale virtual reality training. Action based assessment does not provide a method of capturing a trainee performance, making it impossible for an assessor to directly observe a trainee. As a consequence, the actions conducted by the trainee could be the result of luck rather than logical application of knowledge. Furthermore, because the assessment is unique to a single simulation environment, the action based criteria has to be manually reconfigured by a skilled expert for each simulation, rather than offering a generic approach which adapts to any simulation.

2.7.6 Animation Recordings

Animation recording is a method of capturing simulation data in a format that replays the data in a visual, but not logical format. These recordings are embedded into a simulation environment and operate by capturing actions and modifications of data elements from the trainee performance to an animated file which can be replayed for direct observation.

Lopez et al. (2017) used a technique that creates and stores animation of movement and object manipulations conducted by the trainee, allowing the captured data to be replayed. Although effective, Lopez et al.'s (2017) system is restricted at re-creating the motions of the simulation and does not store low-level information about the interaction. Furthermore, like most approaches for animation, Lopez et al.'s (Lopez et al., 2017) system stores all recorded data into an engine supproted animation file. This means the animation can only be accessed and reproduced within the engine, preventing any analysis of the performance to be conducted in external programs.

In maintenance training Numfu, Riel, and Noël (2019) used a similar animation technique as Lopez et al. (2017), using an animation recording module to capture the posture, gestures and transform of a primitive virtual environment into an animation file. In their case study, Numfu, Riel, and Noël (2019) a trainee operated a training simulation to replacing a saw blade. The trainee tackled the simulation utilising a combination of keyboard and mouse input, along with a leap motion device which provided hand and gesture tracking (Numfu, Riel, and Noël, 2019). The animation can then be replayed to view the performance of the participant or used as a guide to capture a best practice scenario to inform future trainees (Numfu, Riel, and Noël, 2019). However, like Lopez et al.'s (2017) work, Numfu, Riel, and Noël (2019) reproduces the virtual elements of the data, but not the logical modifications. Furthermore, the animation only capture selective elements of the performance, making the assessment rely on observation of the visual reconstruction alone.

Both works by Lopez et al. (2017) and Numfu, Riel, and Noël (2019) are incorporated into the Unity game engine. However, Unity offers its own animation recording feature which is designed to capture gameplay modifications and store them as animations, called GameObject Recorder (Unity, 2021a; Unity, 2021c). The recorder captures a hierarchy objects and their subsequent modifications, allowing their values to be replayed to replay the modifications applied to the object (Unity, 2021a; Unity, 2021c). The intention of this feature is to record physics events as animations, allowing the physic elements to be replayed from an animation rather than impacting the performance at run-time. However, the recording feature only works within the Unity Engine development environment, and it is not supported within individual builds of games or simulations, making the feature irrelevant for the purpose of capturing VR simulations. Furthermore, the recording animation system can experience considerable lag or system crashes when large amounts of data is being recorded, causing responsiveness issues. As a result, the Unity gameobject recorder (Unity, 2021a; Unity, 2021c) and works by Lopez et al. (2017) and Numfu, Riel, and Noël (2019) are incapable for of recording a complete capture of a gameplay or training session.

While animations can be useful aids to visualise the trainee experience, limitations of performance and interaction indicators, along with the abstract format used for the recordings (restricted to engine), make animation recording too restrictive for conducting useful assessment in room-scale virtual reality.

2.7.7 Environment Recording

Environment recordings are forms of recording which capture variable information within a game or simulation, using similar methodology as animation recordings (see Section 2.7.6), but with several key differences. With environment recordings, the focus is capturing sufficient data that would enable relevant sections of the environment to be captured for reproduction (Goldberg, Knerr, and Grosse, 2003; Von Spiczak et al., 2007; Raij and Lok, 2008). Knowing that assessors are capable of conducting manual assessment of trainee performances (Hanoun and Nahavandi, 2018), environmental recordings appear an effective method of visually capturing and reproducing the trainee performance. However, there are many implementations of environment recordings, each designed to suit a unique engine environment or simulation scenario.

One of the earliest examples of capturing simulation environments and experiences is Goldberg, Knerr, and Grosse's (2003) work, which used early implementations of VR technology to replay and review army training simulations in an effort to measure trainees performance. Goldberg, Knerr and Grosse's (Goldberg, Knerr, and Grosse, 2003) after-action review system used event data collection processes to capture the events that took place during a simulation. While this offers a guarantee that an events can be triggered and reviewed repeatedly during the after-action review, the input of the learner is not recorded. Because the focus of the recording is on the event data of the training performance, it is difficult to identify any usability issues experienced by the trainee since their input is not replicated. Furthermore, because one of the key advantages of virtual reality is the ability to deploy the training sessions remotely, understanding trainee usability issues is a key requirement of understanding the trainee performance.

Greenhalgh et al. (Greenhalgh et al., 2002) developed a technique of temporal links which enabled the recording of data in the MASSIVE-3 system (Greenhalgh, Purbrick, and Snowdon, 2000), with the objective to link prior recordings to real-time VR environments. Greenhalgh et al. (2002) system operates by recording all changes made to the virtual environment at run-time. The focus of Greenhalgh et al.'s (2002) work was integrating VR data for media use in film and television and incorporation of previous VR experiences into a real-time VR environment to create new content or review experiences. Using temporal links, Greenhalgh et al.'s (2002) method allowed recorded virtual environment can be replayed and embedded into a live virtual environment for purposes of extending the live virtual environments content or narrative structure. Although not stated, the implementation of temporal links (Greenhalgh et al., 2002) appears to be designed for desktop VR and not head-mounted display VR equipment. Likewise the description of the temporal links system (Greenhalgh et al., 2002) hints that the recording and reconstruction of the temporal links uses animation-like recording of the virtual environment rather than a raw data capture of the VR trainee performance data. In a use-case example for reviewing virtual experiences (Greenhalgh et al., 2002), the author's describe the ability to replay the virtual environment from any perspective but do not discuss any ability for interaction of trainee input or variable monitoring of recorded objects. This lack of clarity suggests the temporal links system is animation based which suffers from the same limitations as other animation reconstruction systems discussed in Section 2.7.6, including the lack of raw user-input data, which limits the understanding of usability issues from input devices, which can be crucial in understanding the context of the trainee in room-scale virtual reality scenarios.

Inspired by Greenhalgh et al.'s (2002) temporal links work, Raij and Lok (2008) created Interpersonal Scenario Visualizer (IPSViz), an after-action review tool for human-virtual human experiences, sharing similarities in reproducing data. Raij and Lok's (2008) tool captures tracking data, audio,video, speech, event logs and human behavior, with the objective of reproducing the simulation experience with a virtual-human, overlaid with additional data (Figure 2.3) On-top of the video and audio data reproduced, 3D reconstructed rendering, sensor data and transcripts of conversations were overlaid to provide in-depth contextual knowledge of the real-world posture and location of the trainee relative to the virtual simulation data. Raij and Lok's (2008) approach is a hybrid model that uses video recordings to replay three different perspectives of the simulation, along with reconstructing the 3D environment for the virtual human. IPSViz is designed to capture a trainee performance to enable self-evaluation and review, enabling the trainees to gain insight into their performance to improve future work (Raij and Lok, 2008). However, IPSViz focuses on communication data rather than reconstruction of sophisticated simulation experiences, as the focus of Raij and Lok's (2008) work is to enable trainees to evaluate their performance during medial patient interview, using human to virtual human experience as the structure of the simulation. Because the simulation environment is minimal, the capture of 3D rendered data does not require significant attention. In comparison to current room-scale virtual reality training simulations which would require significant capture of variable information as necessary to enable authentic assessment of the trainee performance, IPSViz emphasised communication methods such as voice, audio, gaze and kinetic movement. While the IPSViz system enables an effective method of self-review of human to virtual human communication, it is not designed to capture and reconstruct room-scale virtual reality training simulations.

Backlund et al. (2007) created a firefighting simulation as a modification to the game Half Life 2. The simulation was designed to supplement real-world training to determine search strategies and orientation skills in unfamiliar environments (Backlund et al., 2007). Embedded within the simulation was a supervising tool that captured the metadata of the trainee performance during the simulation (Backlund et al., 2007). This incorporation enabled the trainee performance to be replayed exactly as was experienced by the trainee, enabling observation of the performance similar to video, but without having to store any of the visual information (Backlund et al., 2007), reducing the storage size of the data. Furthermore, the tool captures additional gameplay data including the health and transform data of the trainee ten times per second, creating a log of relevant areas of assessment information. The tool can then reconstruct an orthographic view of the training environment, demonstrating where the trainee has been located while exploring the firefighting en-

Figure 2.3: Raij and Lok's (2008) Interpersonal Scenario Visualizer tool (IPSViz) for reviewing virtual patients interactions (Image published with author's permission).

vironment (Backlund et al., 2007). The logging of relevant data at time instances and orthographic movement mapping is a similar implementation to background work discussed in Section 3.6 that inspired this PhD research work. As will be discussed later (see Section 3.6) , the orthographic and time instanced recording data lack sufficient context to the trainee performance, as the information is too limited to acquire in-depth insight into the trainee performance. The work by Backlund et al. (2007) overcame this limitation by using the built-in recording feature of the game engine for Half Life 2, allowing a visual reconstruction of the trainee performance that was supplemented by the logged data. However, the findings from Backlund et al. (2007) suggest that the visual observation of data alone is insufficient or at-least inefficient for judgement of trainee performance, otherwise the reconstruction of trainee performance data from the game-engine would have been sufficient to determine the movement decisions by the trainee. By also recording additional trainee performance data Backlund et al. (2007) assisted the assessment of the trainee data, recording information about the trainee performance that was not automatically captured by the game engine implementation.

Stone, Snell, and Cooke (2016) created 3D underwater mine countermeasure training simulation. To capture trainee data for assessment, the simulation logs trainee data that is time-stamped with their progress exploring the underwater virtual environment, including other search parameters for dwell times and identification of objects, both correct and incorrect (Stone, Snell, and Cooke, 2016). To conduct after-action review, a summary screen presented high-level performance indicators of the trainee, highlighting the summative information for objects identified and hit percentage (Stone, Snell, and Cooke, 2016). After reviewing the summative data, the performance could then be replayed using a non-linear reconstruction methodology which enabled a visual observation of the trainee performance by the assessor (Figure 2.4) (Stone, Snell, and Cooke, 2016). The replay tool included embedded feedback tools which enabled the assessor to highlight on the screen by virtually drawing onto the reconstruction while discussing the feedback with the trainee. Furthermore, the tool includes the ability for the assessor to view the trainee from multiple perspectives, including a free-camera control which offered the assessor full control of the observation and provided an assisted feature that reproduced the search path of the trainee during the dive (Figure 2.5) (Stone, Snell, and Cooke, 2016). However, while the assessment method enables visual reproduction of the trainee performance, the approach requires additional data to be captured, facing the same problem as experienced by Backlund et al. (2007).

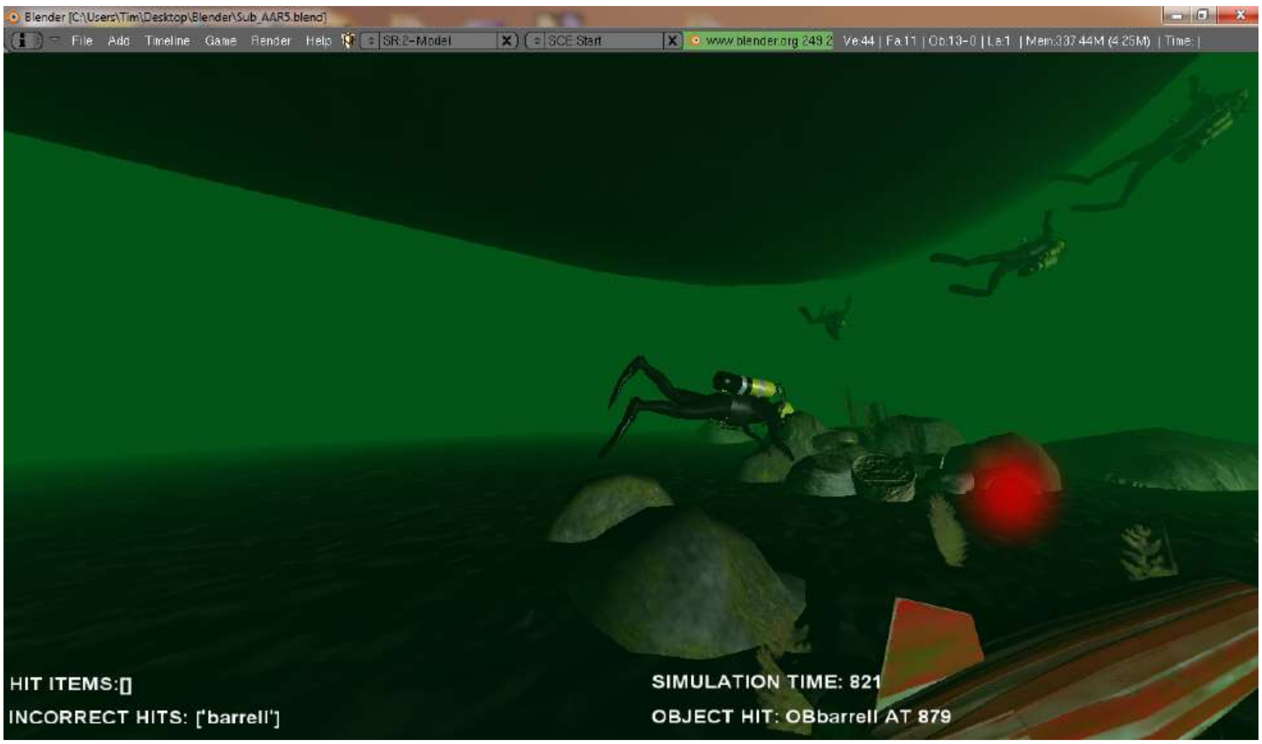


Figure 2.4: Stone, Snell, and Cooke's (2016) After action review screen interface (Image published with author's permission).

Both methods of assessment (Stone, Snell, and Cooke, 2016; Backlund et al., 2007) enabled consistent observation of a trainee performance for after-action review, yet the approach required further information to be captured to enable an effective method of assessment. In the work of Stone, Snell, and Cooke (2016), the additional data was used to provide a summative impression of the trainee performance prior to observation, whereas Backlund et al. (2007) used the data to create a movement map of the trainee exploration. Consequently, both forms of additional data were designed to present an assisted method to assess the trainee performance that could be verified by an assessor. Due to the recording methodology used, additional implementations of recording data were required, signifying that visual reconstruction of the trainee movement and actions alone were insufficient for effective evaluation and assessment. Furthermore, findings from Stone, Snell, and

Figure 2.5: Stone, Snell, and Cooke's (2016) After action review screen demonstrating assessors rear-perspective with a diver search path overlaid onto the assessment environment (Image published with author's permission).

Cooke (2016) and Backlund et al. (2007) build upon the assessment potential of Raij and Lok's (2008) IPSViz system, which captured additional external data such as sensory information of the trainee to improve the data available for assessment. These papers show that it is important to consider that, within simulations, it may be necessary to capture data that is not directly embedded into the simulation outcomes.

The discussed approaches differ based on the requirements of the simulation, which drives the focus of the tool. The discussed works are all designed to capture a specific simulation or resolve a specific gap within a sub-section of research, and are often constrained within that gap. For example, Raij and Lok (2008) design their approach on the focus on communication data of the trainee performance, as that was the objective of the simulation. However, this does not mean the Raij and Lok's (2008) approach would be suitable for room-scale virtual reality training simulations, as the goals and requirements differ. As a consequence, the discussed approaches are difficult to adopt to other scenarios.

What these implementations do provide, is insight into the features they incorporate to improve the overall assessment experience. Raij and Lok's (2008) IPSViz encapsulated simulation recordings with real-world video, transcribed audio and sensory information, and Stone, Snell, and Cooke (2016) overlaid dive-search patterns and incorporated drawing functionality. These features are intended to improve the assessment and feedback potential of the approaches.

In summary, environmental recording improves upon animation based recordings by offering more flexibility in the approach for recording data, demonstrating an effective method of observing training simulations. However, existing implementations are incompatible with the aims of this work, due to their incompatibility with modern simulations (Goldberg, Knerr, and Grosse, 2003; Greenhalgh et al., 2002; Von Spiczak et al., 2007) or constrained recording and reconstruction implementations (Raij and Lok, 2008; Stone, Snell, and Cooke, 2016).

2.7.8 Video Assessment

Unlike all other assessment methods discussed in this thesis, *video assessment* is the only method that is capable of recording data without being directly embedded within the simulation environment. Video and audio recordings are a common way for studies in laboratory and 'in the wild' to document experiments and procedures that can be analysed qualitatively (FitzGerald, 2012). Video recordings make it possible to analyse the the performance of a trainee simulation through direct visual observation, and are stored in a format that is easy to share between collaborators (FitzGerald, 2012).

Video assessment is useful and easy to implement method that allows a recording of a training performance to be captured for assessment by assessors, with non-linear functionality (Roberts et al., 2017). During assessment the video can be paused, replayed, edited and zoomed in and focus on areas for focused observations (Fukkink, Trienekens, and Kramer, 2011). The use of video capture to record VR simulations (Patle et al., 2019; Hanoun and Nahavandi, 2018) is due to the prevalence and accessibility of video recording software and its compatibility (FitzGerald, 2012). Furthermore, video recordings can be used to capture the simulation, and the real-world movement and interaction of the trainee operating the VR equipment (Lazar, Feng, and Hochheiser, 2017).

Because video technology is advanced and well understood, videos can quickly be skimmed through timestamps, which are used to identify and mark points of interest from the trainee performance (Hulsman and Vloodt, 2015). Video assessment also offers the ability extend its capability by incorporate software-based video motion tracking (Ganni et al., 2018). The inclusion of motion tracking software can enhance the assessment capabilities, allowing the motion data to be used to create a set of criteria for videos to be compared against (Ganni et al., 2018). Software can also be used to enrich the data captured, by providing additional information such as the path of projection, speed of movement, timing of tasks and drastic motions made (Ganni et al., 2018). This data is similar to the motion tracking work proposed by MacCormick et al. (2019) discussed in Section 2.7.3, however, with video, the context to the enriched data is available for direct observation.

While most approaches to video assessment only discuss the use of a single video recording. Abdelaal et al. (2020) used two camera to explore the impact of training and assessment in robot-assisted surgery. For assessment, assessor's viewed a single and dual perspective of video recording of the surgical task, with the objective of counting the number of errors committed by the participant in training (Abdelaal et al., 2020). Dual video assessment increased the accuracy by 9% compared to single video (Abdelaal et al., 2020). Abdelaal et al. (2020) mention a future implication of their findings is to provide additional perspectives for virtual reality assessment, with hope it would improve error analysis and skill assessment (Abdelaal et al., 2020). With room-scale virtual reality, one video perspective of a room-scale capable virtual training simulation will be insufficient.

The use of multi-video perspectives is prevalent in multi-user training simulations, Thurston and Martin (2011) demonstrates a multi-viewpoint video after action review system, referred to as GEAARS, captures gameplay of trainees from their perspective, storing multiple videos from each trainee within a game scenario (Thurston and Martin, 2011). For review, GEAARS incorporates all captured videos into a single assessment environment that syncs all the recordings together and presented them within a single observation environment (Figure 2.6) (Thurston and Martin, 2011). To improve the usefulness of the assessment method, GEAARS included an additional top-down recording of the gameplay environment, which provided additional context to the trainee perspective recordings (Thurston and Martin, 2011), offering alternative perspectives of for observation of gameplay. The design of GEEARS suggest that Abdelaal et al.'s (2020) use of multiple perspectives has merit and justification, since it provides a level of redundancy that can improve the reliability of the method for conducting assessment. However, with more perspectives captured to video, the more storage spare required to record and archive the performances.

While video assessment appear to improve upon metric forms of data capture, there are still flaws in the approach. Because video is often captured from the perspective of the VR trainee, the video can not guarantee knowledge regarding the input state or location of tracked devices, and since virtual reality is a room-scale immersive format, it is also easy for participants to accidental or deliberately obscure their perspective when completing actions (Ganni et al., 2018). This means assessors are forced to evaluate work which could not be visually observed due to the obstruction, requiring an approach to determining how to best judge areas of the assessment (Ganni et al., 2018). Even though video is capable of providing a visual context for observation most of the time, remote or distance sessions are likely to be unreliable, given the concerns of video obstruction and lack of input actions. In these situations, the video recording would have to be supplemented with additional software or hardware (Petrie et al., 2006). Furthermore, video observation methods also require the trainee to have access to suitable recording hardware and accompanying software that is capable of recording the training performance. Even if recording hardware and software are available, what is visible from the field of view (FoV) of the trainee in VR is cropped during recording. As a result, the reduced FoV of the mirrored video being recorded results in the peripheral view of the trainee perspective being missed, causing a loss of observation data.

Depending on the type and configuration of training simulations, it is possible that in-game video captures of the trainee would be insufficient to identify constraints in the trainee performance related to room-scale movement. For example, objectives outside the reach of the trainee physical

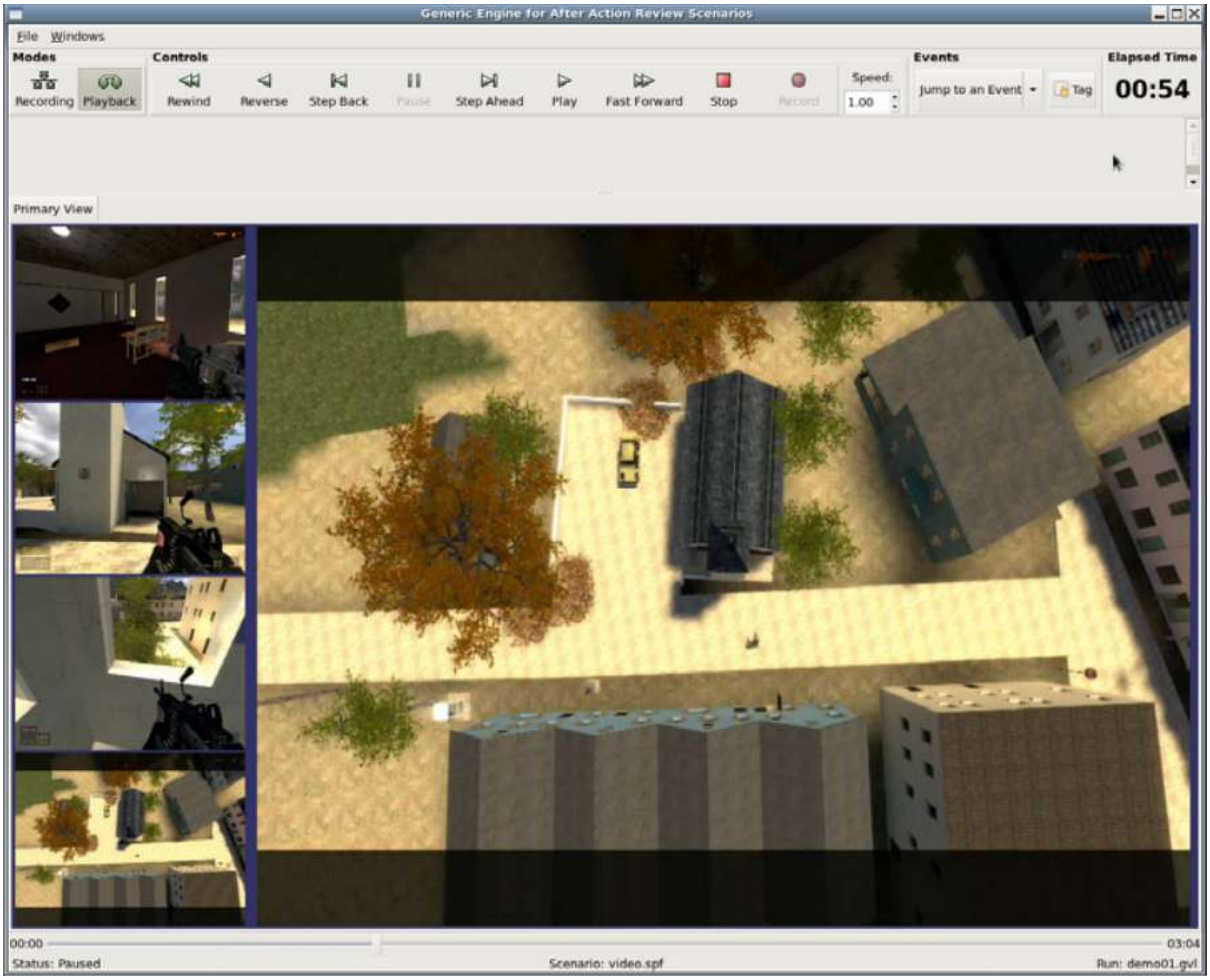


Figure 2.6: Thurston and Martin's (2011) Generic After Action Review System interface (Image published with author's permission).

movement space, resulting in failure due to lack of ability, rather than lack of knowledge (Howie and Gilardi, 2019). A potential solution to this problem is to capture both virtual and real-world videos of the trainee performance for assessment, generating dual context for both environments (Howie and Gilardi, 2019). However, the presence of a video camera in real-world could alter the assessment outcomes, as the recording equipment may cause unnecessary pressure and nervousness on the trainee (Nyström et al., 2014). Face-to-face training would also suffer from the presence of an observer, but it could be argued that it would be less intimidating knowing mistakes or actions could not be replayed or seen by anyone else.

In summary, video assessment improves upon analytical assessment, as the approach is suited for direct observation of the trainee actions. The many features and abilities available make it a robust and effective method of observing trainee performances, with greater reliability when multiple video perspectives are available. However, the predefined nature of video assessments means if data is not captured, missed, or obscured within the video, it lost forever, reducing the ability for reliable assessment. Consequently, video as a method for capturing data limits the quality and quantity of data that can be acquired by the simulation (Grübel et al., 2017). Furthermore, room-scale virtual reality introduces challenges for video assessment which are difficult to overcome, and the storage space required to archive data lineally increases with the amount of video perspectives captured.

2.7.9 Unreal Engine Replay System

The Unreal Game Engine (Epic Games, 2021b), offers functionality to record and reconstruct trainee actions during a gameplay session. This system is referred to as the *Replay System* (Epic Games, 2021a). The Replay System enables gameplay sessions to be captured for trainee observation for the purpose of replaying small gameplay clips, or multiplayer gameplay sessions to the trainee (Epic Games, 2021a). This feature is utilised by games such as Fortnite and Player Unknown Battlegrounds, which both operate using the Unreal Game Engine and capture gameplay during run-time.

Unlike the other environmental recording methods discussed in Section 2.7.7 and in (Howie and Gilardi, 2019), which often required linear reconstruction, Unreal Engine takes a different approach to reproduce non-linear reconstructions of trainee performances. Along with capturing trainee input and non-deterministic data, Unreal Engine enables developers to record additional data which can be used to enable non-linear reconstruction of gameplay (Epic Games, 2021a; Epic Games, 2021b). This means the recording process is designed to capture a large amount of gameplay data rather than just trainee input and non-deterministic data.

By allowing the recording of additional developer defined data, the reconstruction can implement a non-linear reconstruction methodology to reproduce the trainee performance. This means Unreal Engine's implementation records all necessary game data required for rewinding, skipping or fast forwarding. To achieve this, the reconstruction system can use the manually captured data (assigned by a developer) to reconstruct data at any time-stamp instance during the reconstruction process. Therefore, assuming all necessary data is captured, the training simulation can be reconstructed at a specified time instance of the reconstruction without having to linearly go through the process from the beginning (Epic Games, 2021a). However, when incorrectly configured, the reproduced trainee performance can diverge when using any non-linear functionality, as missing data prevents consistent reproduction of the simulation. Furthermore, as the approach only utilises a single layer recording implementation, all data must be captured during run-time. As such, the recording system has to balance saving data without impacting system performance, potentially limiting the amount of data that can be captured.

In normal conditions this would result in the gameplay experienced by the trainee to deteriorate with stuttering and overall poor performance, since the hardware must run the gameplay application alongside the recording of data. Unreal Engine avoids this by setting a limit on the time available for gameplay data to be recorded during a single frame of gameplay. Any remaining data not saved during that single frame is pushed back to the next frame to be recorded. This means that when the rendering time for a frame reaches a defined time, the data not yet saved is pushed back to the next rendering frame of gameplay. This approach minimises the impact on the gameplay experience by prioritising the responsiveness of the game, at the expensive of the accuracy of the recorded data. This causes the recorded data to potentially be out of sync from the actual recorded input, with data potentially being pushed back for recording each frame. As a result, when reproducing the data, there is no guarantee the data will be successfully captured in sync or captured at all, if the recording process continuously exceeds the run-time performance limit.

However, the focus of this recording structure is not designed for or targeting assessment purposes. Unreal Engine uses the Replay System to record gameplay sessions for games, primarily aimed for commercial game experiences, allowing trainees to capture and re-watch their own game sessions or capture small fragments of them for visual observation. In these instances, having data a few frames out of sync is not significantly important. Furthermore, only trainee input and non-deterministic data, including the developer manually configured data is captured during recording. This means that if the recording structure is not correctly configured for recording, any data not captured is lost and cannot be reproduced, potentially impacting the reliability and validity of the trainee performance.

2.7.10 Virtual Reality Assessment Summary

In choosing the approach for conducting assessment in room-scale virtual reality, there are many methods to be considered. Because training simulations are often designed to replicate real-world scenarios, they adopt authentic learning and assessment (Herrington, Reeves, and Oliver, 2007) (see Section 2.4), which provides a logical approach to teaching and training, relying on the trainee to apply their learned knowledge to complete a set of objectives. The methods discussed in this section have varied in approach to conducting assessment, with some methods relying on analytical summative outcomes to determine performance, while others focus on acquiring rich qualitative data for assessment. However, since authentic assessment relies on judgement of trainee logically applying knowledge to complete objectives, analytical indicators only inform assessors if the objective was passed or failed, losing context to the nuances of a trainee performance which demonstrates exactly how a trainee completed their tasks. Furthermore, by ignoring the rich data offered during training performances, opportunities are lost for improvement, since the nuances can provide signals that reveal the true nature of a trainees confidence in their role, such as their body posture or voice (discussed further in Section 4.1.1).

Methods that focused on predefined or automatic outcomes relied on summative indicators for assessing the trainee performance, ignoring the rich data. Whereas, methods that aim to reproduce the trainee performance visually relied on assessors to assess the data. Because relevant literature indicated automatic assessment is not yet capable of replacing the role of assessors (O'Leary et al., 2018; Beller, 2013; Bennett, 2015), human driven assessment appears likely to provide more reliable judgement of trainee competency. The existing automated forms of assessment are forced to dilute the assessment process into predefined parameter conditions, which do not generate confidence or reliability in authentic assessments. This implies that, while prevalent in a wide number of simulations, automated or predefined summative methods of assessment are deemed unsuitable for the purposes of this work.

While summative methods of assessment are unsuitable for conducting authentic assessment, the role of these methods can still play a role in the learning experience, just not as a replacement for a skilled assessor. For example, stealth assessment and action based assessment can be used to provide an overall indication of the trainee performance, which can then be verified by an assessor using an in-depth qualitative visual method. Otherwise, the information of the outcomes can be used to provide assisted assessment aids, as other methods have demonstrated improve usability (Raij and Lok, 2008; Stone, Snell, and Cooke, 2016). However, on their own, these summative methods are considered too unreliable and often will not provide a true representation of the trainee performance, as the learning outcomes are based on predefined parameters and do not attempt to assess or identify the nuances of individual trainees.

Although video should provide a guarantee of accuracy, it lacked reliability for acquiring the information from a trainee performance, with data easily being obscured or not recorded due to the predefined position and perspective of the recording. Therefore, environmental recording and animation based recordings appeared the more robust method of capturing trainee performance data, as it theoretically allows the simulation to be reviewed from any perspective, ignoring any obstructions of data. In comparison, video required multiple cameras to record each different perspective as seen in the works of Thurston and Martin (2011) and Abdelaal et al. (2020). However, this does not mean that environmental and animation based recordings share the same consistency and accuracy as is guaranteed by video. For one, most animation recordings are only capable of capturing and replaying the trainee experience in a restrictive manner using limited data, as demonstrated by the works of Lopez et al. (2017), Numfu, Riel, and Noël (2019) and Unity Engines official implementation (Unity, 2021a; Unity, 2021c). In these works, the animation only captures trainee data, and ignores the wider context of the environment. This is because animations are incapable of recording a large amount of data due to their format, as such, the method is forced to capture only limited amounts of the trainee performance (discussed further in Section 3.8.1), making it unsuitable and unreliable as an approach to capturing data for authentic assessment, leaving environmental recordings.

Environmental recordings traditionally face the same recording issues as animation methods when using conventional recording methodologies. However, as environmental recording technologies are often original development implementations, rather than relying on existing animation architecture, it is possible to develop alternative methods of capturing data. The recording methods identified in the literature were designed to focus on trainee oriented data only (Raij and Lok, 2008; Backlund et al., 2007; Stone, Snell, and Cooke, 2016), with no identified implementation demonstrating the ability to capture all data within a simulation. Consequently, the lack of access to all data impacts the assessment potential for several methods, causing them to rely on the limited and predefined data captured, potentially losing context to the trainee performance.

When discussing environment recording implementations, it is important to clarify that not all existing implementations are designed with the intention of providing assessment (Epic Games, 2021b; Greenhalgh et al., 2002). For example, the work by Raij and Lok's (2008) was designed to provide an assessment system to evaluate a virtual and human participant in a medical setting, as a result, the tool included many assisted features such as transcribed audio and human behaviour logs which is overlaid and synced to the visual observation which enabled an effective method of providing assessment. These tools offer advantages over other implementations of assessment, as they improve and assist the assessment experience (Raij and Lok, 2008). Stone, Snell, and Cooke (2016) presented summative indicators of trainee performance prior to observation and included assisted features within the assessment process, including an overlay of the trainee dive search path. Furthermore, the tool also included built-in feedback options that enabled drawing on the screen, enabling assessors to point and highlight areas of concern to a trainee during subsequent discussions.

In order to fulfil these objectives, the framework must be designed to focus on the overall experience for capturing and reproducing data in a format that is useful and usable for assessors to conduct authentic assessment and generate contextual feedback. To achieve this it is important to understand how feedback should be incorporated into this process, and identify what options are available and suitable for supplementing virtual forms of assessment.

2.8 Feedback of Virtual Reality Training Simulations

Feedback that derives from a external source can provide the maximum learning benefit (Raymond and Usherwood, 2013). With formative feedback, assessors provide trainees with information gathered during assessment that informs them of future work for improvement or correcting mistakes and misunderstandings (Shute, 2009). As discussed in Section 2.3.2, feedback is a critical component in the learning module of training (Wijewickrema et al., 2017) which should provide constructive feedback that corrects mistakes and praises positive actions of trainees. It is important that this feedback is provided in a timely manner (Wijewickrema et al., 2017) and sensibly delivered, to avoid trainees considering negative feedback as a form of criticism and a sign of weakness (Broquet and Dewan, 2016; Broquet and Punwani, 2012). Ideally, the ultimate goal of VR training should be autonomous assessment and feedback, but as discussed in Section 2.7, that is not yet possible and assessment and feedback still relies on human assessors.

2.8.1 Written Feedback

Written feedback is universally used within education, however, it is important to consider that the quality of comments provided by an expert using written feedback can be reliant on their willingness, time available and ability to condense thoughts into sentences that are informative and usable by a trainee (Sadler, 1989). In some cases, the written feedback may not be easily digestible, since it may lack reference to the highlighted flaws in assessed work (Sadler, 1989) or it may be difficult to determine meaning and tone behind comments and guidance written by the assessor. In these situations, incorrect understanding of feedback can nullify opportunities that could have been used to allow a trainee to identify and rectify their mistakes and misunderstandings (Deeley, 2018). In virtual reality simulations, the conventional approach used in education for written summaries and overlaid reports may be incompatible due to the medium differences, with virtual reality simulations relying on the context and authenticity of the logical application of knowledge. Because the medium is visually driven, it is more difficult to write written feedback that is detached from the simulation context. As such, it is unlikely that standalone written feedback can be consistently written to a high standard, without information being lost, causing a deterioration of the feedback quality (Sadler, 1989; Deeley, 2018). Therefore, unless written feedback is contextualised in a format that is overlaid onto the virtual context, it is unlikely written feedback is the correct approach for providing feedback in VR.

2.8.2 Video Feedback

In contrast to written feedback, video feedback provides a visual context and often along with verbal narration, which contains the information the feedback content. While it is too early to determine the effect video feedback has on learning outcomes (Mahoney, Macfarlane, and Ajjawi, 2019), its use has been advocated (Mahoney, Macfarlane, and Ajjawi, 2019), with most studies demonstrating improved or equal video feedback preference over written feedback (Kruger and Sage, 2020; Lamey, 2015; Matthews, 2019).

A study conducted by Lamey (2015) found that the majority of trainees had positive impressions of video feedback, and ranked it higher than traditional written feedback. Likewise, a study byDeeley (2018) found video feedback to be popular with trainees in their cohort, with some trainees watching the video feedback was like having an expert present going through the paper, providing feedback on the good and bad parts of their work. Along with providing corrective feedback, positive feedback can result in improvements for trainee performance and acquisition of motor skills (El Boghdady and Alijani, 2017). Furthermore, Deeley (2018) stated that using video feedback resulted in greater amounts of feedback, as well as indicating areas the expert found concerning (Deeley, 2018). These studies suggest that feedback when supplied using a video format results in more content and context, improving the overall quality of the feedback.

Video makes it easy to provide personalised feedback (Mahoney, Macfarlane, and Ajjawi, 2019;

Kruger and Sage, 2020), which reduces the time required to generate (Lamey, 2015) and provides the ability to target individual flaws and offer constructive guidance for improvement (Lamey, 2015; Kruger and Sage, 2020). Trainee's often preferred video feedback as it provided detailed information and the tone of the assessors voice could be used to gauge the importance of comments (Lamey, 2015; Kruger and Sage, 2020). The process of providing video feedback was also shown to create trust between trainees and assessors (Kruger and Sage, 2020; Matthews, 2019), and enabled trainees to easily revisit feedback with context prior to future assessments (Matthews, 2019).

Lamey (2015) shows that for video feedback to be effective it must be detailed and specific to the trainee's work. As a consequence, if video feedback was to be utilised within virtual training simulations, the feedback would need to be isolated to target the individual sections that provide context to the source of the feedback. Furthermore, because VR is a visual driven medium, the inclusion of trainee context is critical to linking feedback to participant actions during a simulation. Lamey (2015) highlight the important differences between written and video feedback, stating that "Written feedback is directed towards the assignment" (Lamey, 2015, p. 701) whereas "Video feedback, in contrast, is directed toward to the trainee" (Lamey, 2015, p. 701). As a consequence, Fukkink, Trienekens, and Kramer (2011) found that observation of video feedback was more effective than other feedback models for explanations, modeling and practice, with recommendations that video feedback is implemented in future research work to improve the feedback given to trainees.

Another benefit to video feedback is the ability for self-evaluation and peer-feedback, with comments and annotations added to relevant instances of a video timeline (Hulsman and Vloodt, 2015). By using video to provide feedback, it is easier to highlight and mark relevant areas that feedback focuses on. Furthermore, video feedback contextualises the information provided by a peer or assessor. Consequently, when reviewing videos with an assessor present, the trainee is able to ask them questions or express concerns immediately (Nyström et al., 2014). However, reviewing feedback at the same time as an assessor could cause tension and anxiety for the trainee (Nyström et al., 2014).

A useful process for video feedback is the ability to replay the video to individuals or teams, while an assessors provides feedback and addresses gaps in participant knowledge (Oseni et al., 2017; Nyström et al., 2014). Videos are also portable, so they can be played on a large amount of devices without an expert needing to be present (Deeley, 2018). In group activities the use of video can provide a method of facilitating discussion between peers to improve the confidence, communication and quality of teamwork (Oseni et al., 2017; Nyström et al., 2014). Self-reflection via replaying videos can also improve trainee knowledge, allowing them to review and reflect on their behaviour and actions during a training performance (Fukkink, Trienekens, and Kramer, 2011; Nyström et al., 2014).

Although video feedback has not been proven to improve learning outcomes and further large scale studies are required to compare it to other forms of feedback (Mahoney, Macfarlane, and

Ajjawi, 2019), the general findings in literature is that its usage has a positive impression on both trainees and assessors (Mahoney, Macfarlane, and Ajjawi, 2019; Matthews, 2019; Nyström et al., 2014). The noticeable drawbacks are the time required setup and use the format to conduct feedback (Mahoney, Macfarlane, and Ajjawi, 2019; Matthews, 2019; El Boghdady and Alijani, 2017). Furthermore, having to record multiple iterations of a video due to mistakes or errors of feedback can greatly impact the efficiency of the process (Lamey, 2015), and should be considered a concern and potential cause of irritation for assessors. Also, because video feedback can require significant storage space and bandwidth allocation for transferring the video (Thurston and Martin, 2011), the process could be expensive and time-consuming. Most prominently, while several studies have advocated or promoted the use of video feedback (Mahoney, Macfarlane, and Ajjawi, 2019; Lamey, 2015; Matthews, 2019), it can be presumed that video feedback is not often conducted in real-world training, as replaying videos for feedback may cause stress and anxiety in participants who have difficulty re-watching themselves (Nyström et al., 2014).

2.8.3 Automatic Feedback

Automatic feedback, requires automatic assessment, and are designed to use simulator analytics to determine the outcome of the supplied answer or action against expected criteria, generating an estimated accuracy of performance (Konakondla, Fong, and Schirmer, 2017). In dynamic analytical data assessment, there are seven analytical data that can be used to measure trainee performance: evaluation subject, evaluation object, purpose, metrics, method, result and time (Pan et al., 2015). Since simulations can operate using cause and effect relationships (see Section 2.7.5) or be solved using analytics strategies (see Section 2.7.2), it is possible to create systems that provide automatic feedback (Konakondla, Fong, and Schirmer, 2017; O'Leary et al., 2018) using these strategies to determine the assessment outcome. These values can provide an objective measurement of skills that is not impacted by assessor bias (Thomsen et al., 2017), assuming adequate assessment and weighting of values is achieved by the analytics that generate a true presentation of the learning outcomes.

However, while an estimated assessment of a performance can be acquired using automated strategies (see Section 2.7), automatic assessment is not yet capable of determining learning outcomes in complex skill-sets or rich data (O'Leary et al., 2018; Bennett, 2015; Armas, Tori, and Netto, 2020). Most automated assessment methods rely on analytic data and are structured and rigid, forcing the data to be evaluated against predefined learning outcomes (O'Leary et al., 2018). Therefore, automatic feedback can only be summative, since analytics values either are or are not within the range of valid results for success (Goderstad et al., 2019; Nicol and Macfarlane-Dick, 2006), and with the role of feedback being to improve trainee misunderstanding of skills (Broquet and Dewan, 2016; Sadler, 1989), poor quality feedback can have negligible value for the learner (Deeley, 2018), making automatic feedback unsuitable for authentic assessment.

However, that does not mean automatic assessment does not have any role within training simulations. Hanoun and Nahavandi (2018) argue that a hybrid approach may be best suited to combine assessor feedback and automated assessment feedback, minimising administrative work, while allowing detailed discussions to take place with an assessor.

2.9 Chapter Summary and Discussion

Authentic assessment has been identified and discussed, presenting its role in the judgement of trainee competency. With authentic assessment, the core principle is to judge trainee performance against the logical application of knowledge that represents the process of real-world situation. However, for authentic assessment to be conducted, trainee performances must be directly observed to determine how the trainee has performed within the simulation, evaluating their application of knowledge and identifying gaps that require intervention by assessors.

This chapter highlighted a gap in virtual reality training assessment, discovering that there is no frameworks suitable for conducting authentic assessment and generating contextual feedback of room-scale VR training simulations. This work aims to contribute to this gap by introducing a novel framework. Several methods have been identified in the literature for conducting assessment in virtual training simulations, including: analytical data, motion and input tracking, video, animations, environment recording, action based and stealth-assessment, but none of these were found to be suitable for adoption into the framework proposed in this work

As discussed, automatic assessment is not yet capable of replacing the role of assessors for conducting after-action assessment (O'Leary et al., 2018; Beller, 2013; Bennett, 2015). As such, authentic assessment of VR trainee performances must be evaluated by assessors. To achieve this, trainee performances must be captured in a format that enables direct observation by assessors, which is necessary for measuring authentic assessment objectives and providing contextual feedback that is meaningful and improves learning (Martins and Kellermanns, 2004; Sadler, 1989; Ellaway and Masters, 2008). In classroom-based training, direct observation of a trainee enables assessors to determine via visual and audio cues how the trainee is coping with the tasks and objectives (Wiggins, 1990). These observations are used to direct the assessor's intervention if a trainee is struggling or requires further assistance. However, the configuration of VR training makes direct observation of trainee's difficult.

With room-scale VR technology is becoming widely adopted (Moore et al., 2020; Berndt et al., 2018; Shaw et al., 2019; Bayerl et al., 2019), thanks to the advances in consumer VR equipment (HTC Vive, 2019), a framework should consider the implications of room-scale capabilities within simulations when capturing, replaying or assessing trainee performances. As a consequence of room-scale VR, predefined methods of observation like video recording are unreliable, as the method would require multiple recordings from different perspectives of the trainee performance to capture all potential trainee interaction.

The literature review demonstrates the potential for recording and replaying virtual experiences and simulations, but also highlight the limitations currently faced. Notably, existing methods suffered from an impact of poor frame rate when recording lots of data (see Section 2.7.6), along with general concerns of captured data being unreliable due to: obstruction of data (see Section 2.7.8), lack of context for the data (see Section 2.7.3) and inability to observe trainee (see Section 2.7.2). With these considerations, a shift to virtual methods of capturing VR training simulations is required.

Findings from the literature have highlighted the importance of assisted assessment features within the observation, assessment and feedback operation conducted by an assessor. Work by Raij and Lok (2008) and Stone, Snell, and Cooke (2016), demonstrated how additional information of trainee performance could be overlaid onto the observation and assessment to enhance the capabilities of the assessor. With Hanoun and Nahavandi (2018) stating that the role of an after-action assessment framework should "augment and assist the [assessor] to do a better job and relieve their cogitative load" (Hanoun and Nahavandi, 2018, p.6), it is important that the framework incorporates the assisted assessment features.

To provide the functionality for assisted assessment features, the framework needs access to a large amount of trainee performance data. As a consequence, it is important that recording operations are capable of capturing all data from within a simulation. With this approach, no simulation data is lost, and the capabilities of assisted assessment features can be adapted to suit the needs of assessor, as seen in (Raij and Lok, 2008), without being restricted to a limited amount of predefined simulation data. Access to all data from a trainee performance enables in-depth analysis that can assist the observation and assessment process for the assessor. However, existing methods of capturing data within simulations are unsuitable (see Sections 2.7.6, 2.7.7 and 2.9), and an alternative recording methodology is necessary to capture all data within a simulation without impacting the run-time performance.

Finally, as feedback is a critically important element of authentic assessment, the framework should incorporate feedback tools and features that can be used by assessors to generate contextual feedback. The framework must be able to generate feedback that contains sufficient context which identifies trainees misunderstandings and mistakes, while still relevant to the overall outcome of the assessment. As discussed in Section 2.8, it is crucial that feedback is supplied in a timely manner, minimising the delay between when the trainee conducted the assessment and when feedback is generated. As AI technology is not yet at the stage were automatic feedback can replicate or replace a human assessor, assessors are still expected to have a considerable role in providing feedback (Roy, Bakr, and George, 2017). Written feedback is not suitable as it is time inefficient and resource heavy. Video feedback on the other hand, is easy to implement within VR, as it is a visual and verbal driven medium. With video feedback, assessor could provide narrative commentary with visual reference and contextualise to the trainee problem and feedback. To streamline the generation of feedback, the framework should embed the assessment and feedback tools and features within a single evaluation environment.

Taking the literature review findings into consideration, the novel framework introduced in this work must be able to overcome the following requirements and limitations imposed by the existing methods of capturing and replaying virtual simulation data:

1. A recording methodology must be able to capture all the information from a trainee performance, without impacting the run-time performance of the simulation.

2. Assisted assessment features should which provide in-depth analysis and information regarding the trainee performance.

3. Feedback tools that generate contextual feedback should be embedded within the framework to streamline feedback from within the assessment environment.

4. Linear forms of assessment are too time consuming and restricting, therefore the framework must provide a non-linear observation of the experience, providing assessors the ability to rewind or fast-forward as is possible with video methods of observation and assessment.

Findings from the literature review have answered RQ1 and RQ2 of this work, identifying that existing methods of conducting assessment and generating feedback are inadequate, and there is no suitable framework for assessors that enables them to conduct authentic assessment and generate contextual feedback of room-scale VR training simulations. However, the findings from relevant research (Raij and Lok, 2008; Stone, Snell, and Cooke, 2016; Hanoun and Nahavandi, 2018) provide directions and inspiration for what is required of a framework to close this gap in knowledge.

The methodology is presented in the following chapter (Chapter 3), sets out the research aim, detailing the problems which formulate the research questions of this thesis. The methodology for tackling these research questions is presented, discussing the approaches adopted, agile development of the novel framework, and the methodology for sampling participants, conducting analysis and evaluating the framework with assessors from industry and academia .

Chapter 3

Methodology

This chapter discusses the aim of this work (see Section 3.1) and presents the research questions (see Section 3.2), introducing the methodology that form the approach taken. Section 3.3 outlines the thematic method adopted to form the qualitative methodology which answered the research questions in this work. Section 3.4 follows discussing the participant sample, with Section 3.5 discussing the methods taken for answering each individual research question. The unpublished and unsuccessful study which inspired this research is discussed in Section 3.6, which highlights the need for an assessment framework for room-scale VR training simulations. Section 3.7 discusses the agile approach taken for the design and development of the framework. Next, Section 3.8 provides the justification for the research approach. Section 3.9 outlines the steps taken to ensure the research and studies were conducted ethically. Finally, Section 3.10 concludes the chapter, summarising the proposed approach for answering the research questions of this work.

3.1 Aim and Objectives

The aim of the research presented in this thesis was to continue filling the gap highlighted in Chapter 2, by creating a novel framework that provides assessors with the means to monitor, observe and assess virtual trainee performances for authentic assessment, while providing tools that allows the assessor to generate contextual feedback of room-scale VR training simulations. This research was conducted from the perspective of assessors. To understand and answer this problem, three objectives were defined.

The first objective (O1) of this research was to understand the how authentic assessment and contextual feedback is currently performed in virtual training simulations, identifying existing methods usefulness and usability. These findings were supplemented with a study involving assessors who operate classroom-based training in the oil and gas industry. These findings formed a basis of the requirements necessary for the framework proposed in this work.

The second objective (O2) of this work was to design the framework, which enabled a streamlined approach for recording and reconstructing virtual training simulations in a format that is useful and usable for assessors to conduct authentic assessment and generate contextual feedback. To address this, the findings from the first objective influence the framework design and architecture, including its components, tool and features.

Finally, the the last objective (O3) was to evaluate the framework and compare it with existing methods identified in the literature. To address this objective, an implementation of the framework was developed, allowing assessors from industry and academia to evaluate the frameworks suitability and compare it against existing methods. This evaluation was conducted to discover the usability and usefulness of the framework against existing methods, generating in-depth findings from the perspective of assessors.

These objectives were divided into seven research questions that formulated the structure and the approach taken for this work.

3.2 Research Questions

To provide a coherent structure to the research for the objectives presented in Section 3.1, the following seven supporting research questions (RQ) were developed. These research questions were provided to focus the narrative and direction for each objective of this work and are intended to highlight the approach taken for answering each objective. As such, O1 includes two research questions, RQ1 and RQ2, which provide clarity for the literature review searches:

RQ1: What is the state of the art for authentic assessment and contextual feedback of room-scale virtual reality training simulations?
The first research question sought to understand what the state-of-the-art was for authentic assessment and contextual feedback in virtual training simulations. This question, answered in Chapter 2, investigated existing literature to determine what methods exist that enable assessors to conduct authentic assessment and provide feedback of virtual reality room-scale training simulations.

RQ2: What evidence exists on usability and usefulness of identified approaches?
Following the discovery of the methods in RQ1, the second research question attempted to determine the usability and usefulness of the methods discovered. This question focused on identifying the strengths and weakness of the methods based on findings from the literature, and was answered in Chapter 2.

In order to answer O2, three research questions were formed, RQ3, RQ4 and RQ5, which focused on the design of the framework, starting with collecting data as part the requirement study, interpreting the data and finally, the design of the framework components, tools, features:

RQ3: What do assessors need for assessing and generating feedback for room-scale vir-

tual reality training simulations?
To answer the third research question, a study was conducted with assessors from industry to acquire first-hand information of industrial classroom-based assessment. The qualitative study conducted field-observations of an industrial training centre and semi-structured field-interviews with industrial training assessors. This question focused on identifying the requirements the framework and its components, tools and features from the perspective of assessors.

RQ4: What tools are suitable to assess trainee performances in room-scale virtual reality training simulations?
The fourth research question determined what tools and features would be required in a framework for conducting authentic assessment and generating contextual feedback for VR trainee performances. This question was answered by analysing the findings from RQ3 to determine how the requirements identified from industrial training assessors study can be defined into implementable tools and features within the framework.

RQ5: How can findings be compiled to create a framework for a generalised system for observation, assessment and feedback of room-scale virtual reality training simulations?
The fifth research question uses a combination of findings from all previous research questions to generate knowledge for how a framework can be designed and developed for authentic assessment and contextual feedback of room-scale virtual reality training simulations. This question is answered by documenting the design of a framework and all if its accompanying components that enables assessors to conduct authentic assessment and generate contextual feedback of room-scale VR training simulations.

Finally, for answering O3, two final research questions, RQ6 and RQ7, were formed to structure the evaluation of the framework and the comparison against state-of-the-art methods identified within the literature review:

RQ6: Are the tools and features provided in the framework suitable for conducting authentic assessment and generating contextual feedback of room-scale virtual reality training simulations?
Following from the design of the framework proposed in this work in RQ5, research question six evaluates how useful and usable the framework is for conducting authentic assessment and generating contextual feedback of room-scale virtual reality training simulations. A qualitative evaluation was conducted with assessors from industry and academia. The question contributes to knowledge by generating findings for the design of the framework proposed in this work.

RQ7: How useful is the framework compared to state-of-the-art methods for conducting authentic assessment and generating contextual feedback of room-scale virtual reality train-

ing simulations?
The final research question was conducted simultaneous to RQ6, by evaluating the framework against existing methods identified within the literature review in RQ1. Three methods including the framework designed from RQ5, analytical data and video assessment, were evaluated by assessors to determine each methods usefulness and usability for conducting authentic assessment and generating contextual feedback of virtual reality training simulations. The study utilised qualitative approaches for acquiring data, including observations of the methods used, followed by semi-structured interviews. Thematic analysis was conducted to generate findings from the data. These qualitative approaches generate in-depth information of the participants experience conducting the authentic assessment and generating contextual feedback. The question determined the strengths and weaknesses of each method evaluated.

Throughout this work these research questions will be referenced to demonstrate the approach used to answer the objectives (O1, O2, and O3) of this work presented in Section 3.1.

3.3 Research Approach

Because the objective of this study is to discover, interpret and explain authentic assessment and contextual feedback approaches for room-scale VR from the perspective of assessors, a qualitative approach was chosen (Hammarberg, Kirkman, and Lacey, 2016) as the best approach to answer the research questions presented in Section 3.2.

Qualitative research provides the opportunity to dissect a phenomenon by reinterpreting findings in context of the study and real-world (Fossey et al., 2002). Furthermore, qualitative studies emphasises in-depth analysis of the findings in an attempt to contextualise the relationship of the data (Yin, 2015), the findings from studies with assessors can generate the necessary information required to address and answer the research questions. Hammarberg, Kirkman, and Lacey (2016) also note that qualitative methods can be used to analyse participants perspective of their experience, which is not often identifiable or measurable using other approaches. By approaching this research using a qualitative methodology, the findings are sourced directly from the targeted population, providing validity and reliability to the research findings of this work. As such, a contextual design approach was chosen as the qualitative methodology for this research (Holtzblatt, Wendell, and Wood, 2004), involving field studies to generate requirements from end-users (assessors) that are used to design and develop the framework proposed in this work.

However, it is important to consider the drawbacks to this approach. Because the content interpreted and presented while using a qualitative methodology is at the discretion of the researcher, it is important to emphasize that data is analysed, judged and displayed in a trustworthy practice of good conduct (Fossey et al., 2002). To minimise bias using qualitative approaches, the data has been analysed critically, following Braun and Clarke (2019)'s (Braun and Clarke, 2019) guidelines and recommendations to ensure sufficient depth is applied to interpretations of the data, validating the findings are within the context of the data they are used to represent (Sandelowski, 1986). Furthermore, by taking **doi:10.1080/2159676X.2019.170484**'s and Sandelowski (1986)'s advice, the author of this work has attempted to minimise bias by using audio and video recording of the participants dialogue and actions during their involvement with the study to prevent any manipulation of the data, ensuing that the data presented in this work is ethical and suitable for the findings presented.

Because the research questions require studies to be conducted with assessors, semi-structured interviews and field-observations were chosen as the methods for collecting data. With semi-structured interviews, the findings are captured from the thoughts, opinions and voice of the assessors, providing the answers for the research questions from their perspective. Likewise, field-observations enable findings from classroom-based assessment to be studied and used to influence the design of the framework. For RQ3, industrial training assessors were recruited, with the study conducted on-site at the industrial training facility. RQ6 and RQ7 recruited assessors from industry and academia, and was conducted remotely due to COVID19. The findings from these studies with assessors provided in-depth and rich data which required use of a qualitative method for analysis.

3.4 Participant Sample

Assessors within academia and industry were the target audience of this work. Because the concept of this work was to identify the findings for a narrow subject matter, i.e. identifying the design requirements of a framework for conducting authentic assessment and contextual feedback, and then later evaluating it against other methods, the sample size was determined by Malterud, Siersma, and Guassora (2016)'s concept of "information power", which was evaluated by the aim of the study, sample specificity, use of established theory and quality of dialogue. To answer RQ3 and RQ4 (see Chapter 4), and RQ6 and RQ7 (see Chapter 8), a small but targeted sample of participants of five (RQ3 and RQ4) and thirteen (RQ6 and RQ7) provided comprehensive data capturing participants experiences and knowledge (Malterud, Siersma, and Guassora, 2016). Figure 3.1 shows the recruitment of participants for answering RQ3, RQ4, RQ6 and RQ7.

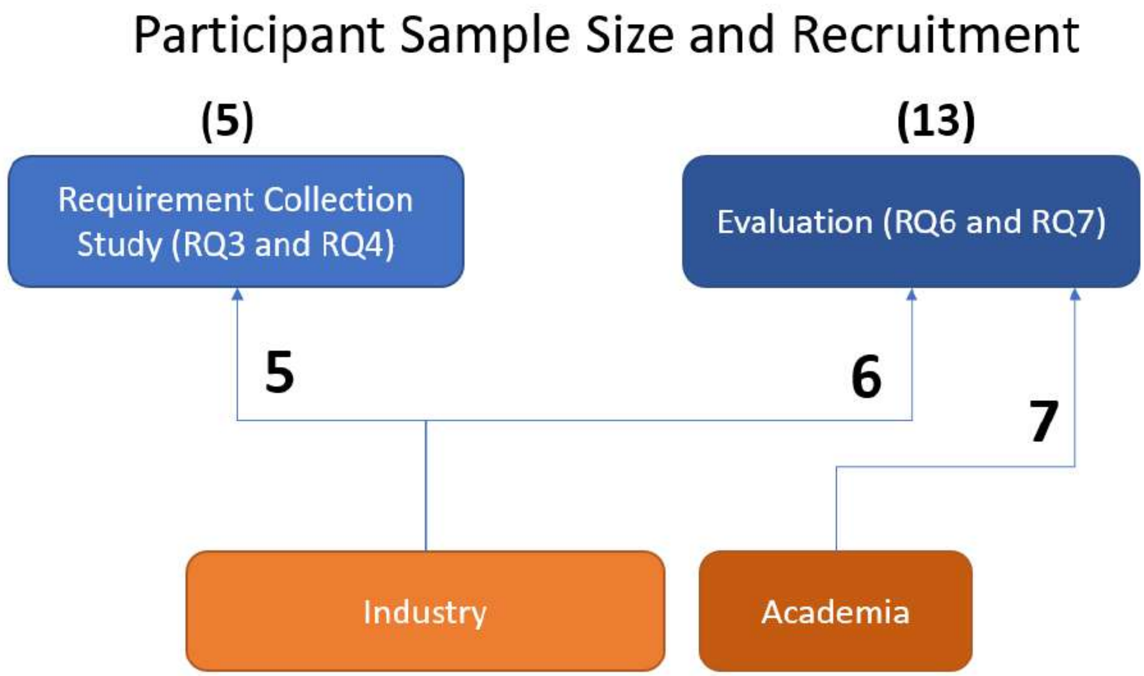


Figure 3.1: A total of five participants from industry were recruited for RQ3 and RQ4. For RQ6 and RQ7, a total of thirteen participants, six from industry and seven from academia were recruited.

3.5 Methods Selected for Answering Research Questions

To create the novel framework that enable assessors to conduct authentic assessment and generate contextual feedback for VR training simulation performances, four elements needed to be taken into account:

- Only a limited number of assessors were available to participate.
- The components, tools, and features required within a framework for authentic assessment and contextual feedback were unknown.
- The findings required open-ended approaches to answering the research questions.
- Usefulness and usability measurement were required to evaluate the framework.

The qualitative approach used for this study put emphasis on the discussions with assessors, generating requirements from their perspective within the focused subject area, providing insights into their thoughts, opinions, and ideas. Discoveries made during qualitative interviews and field observations with assessors for RQ3 provided the requirements for the design and implementation of the framework and its accompanying tools and features for authentic assessment and contextual feedback. Likewise, observations and interviews with assessors in academia and industry (RQ6 and RQ7) enabled the evaluation of the framework proposed in this work. The framework was evaluated against two existing methods for assessment of trainees in VR simulations, allowing identification of themes related to usability and usefulness of the methods using thematic analysis (Braun and Clarke, 2006).

3.5.1 Thematic Analysis Method

Thematic analysis was chosen as the method to identify findings from the data acquired in the study (RQ3) and evaluation (RQ6 and RQ7). Thematic analysis at its core is "a method for identifying, analysing and reporting patterns within data" (Braun and Clarke, 2006) (p.79). Braun and Clarke (2006) recommend six phases, which were followed to analyse the data. Figure 3.2 shows how Brawn and Clarke steps were used to analyse the data collected from the evaluation to answer RQ6 and RQ7. To better contextualise findings, coded sections of the transcripts were linked to segments of video captured during the evaluation which show actions and interactions of the participant while using the methods.

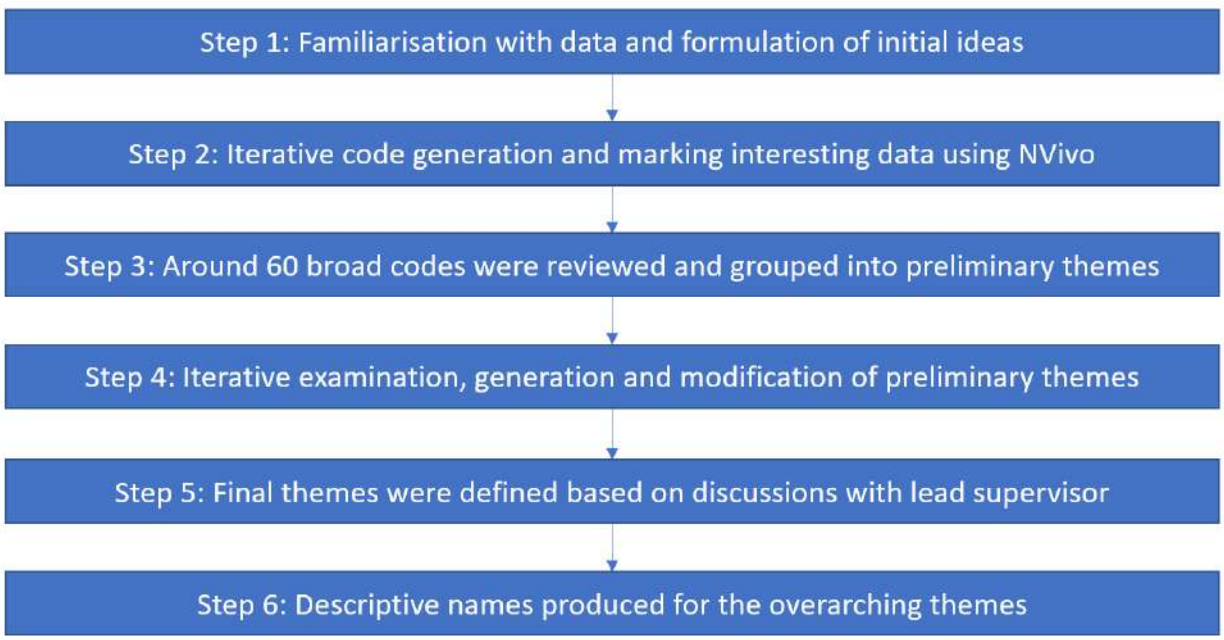


Figure 3.2: A step-by-step account for each section of the analysis process, following Braun and Clarke's (2006) guidelines for answering RQ6 and RQ7 (see Section 8.2).

3.5.2 Research Question 1 and 2

RQ1: *What is the state of the art for authentic assessment and contextual feedback of room-scale virtual reality training simulations?*

To answer this research question, an archival methodology was conducted to search databases and identify relevant literature (see Chapter 2). The methodology used to answer RQ1 is presented in Section 2.1.

RQ2: *What evidence exists on usability and usefulness of identified approaches?*

The second research question examined the strengths and weaknesses of existing methods usability and usefulness for authentic assessment and contextual feedback within training, education and e-learning (see Section 2). This research question is answered using the same methodology as RQ1 (see Section 2.1), utilising the same search terms and databases, since the identification and evaluation of the methods were in overlapping fields.

3.5.3 Research Question 3, 4 and 5

RQ3: *What do assessors need for assessing and generating feedback for room-scale virtual reality training simulations?*

The third research question focused on the acquisition of rich data from assessors using field observations and semi-structured interviews. To answer this research question, it was necessary to use a qualitative methodology as the questions and data acquired is open-ended (see Chapter 4). Five assessors from the oil and gas training industry were recruited to answer RQ 3, using the targeted approach described in Section 3.4. It was important that assessors were recruited from industry which use authentic assessment in classroom-based training and assessment. As such, the five assessors recruited for this study are skilled experts in conducting authentic assessment and generating contextual feedback of trainees performances during classroom-based assessments.

With the in-depth rich data generated from field-observations and semi-structured interviews, a qualitative methodology enabled thematic analysis to be conducted, generating themes from the data which influenced the design of the framework and its components, tools and features by identifying industry training assessors requirements (see Section 4.2).

RQ4: *What tools are suitable to assess trainee performances in room-scale virtual reality training simulations?*

The fourth research question uses the findings from the requirement study with assessors from RQ3 to generate implementable tools and features for the framework (see Section 4.3).

RQ5: *How can findings be compiled to create a framework for a generalised system for observation, assessment and feedback of room-scale virtual reality training simulations?*

The fifth question uses findings from existing research questions to form the design of the framework proposed in this work. By utilising the findings from the literature review, supplemented by interviews and field observations with industrial training assessors. The components, tools and features of the framework were formed, shaping the overall design of the framework that enabled recording, reconstruction, observation, authentic assessment and contextual feedback of virtual reality training performances. This work is presented in Chapters 5 and 6.

Figure 3.3 presents the process used to answer RQ3, RQ4 and RQ5, taking thoughts and ideas from assessors and incorporating them into requirements that form the design of the tools and features within the framework.

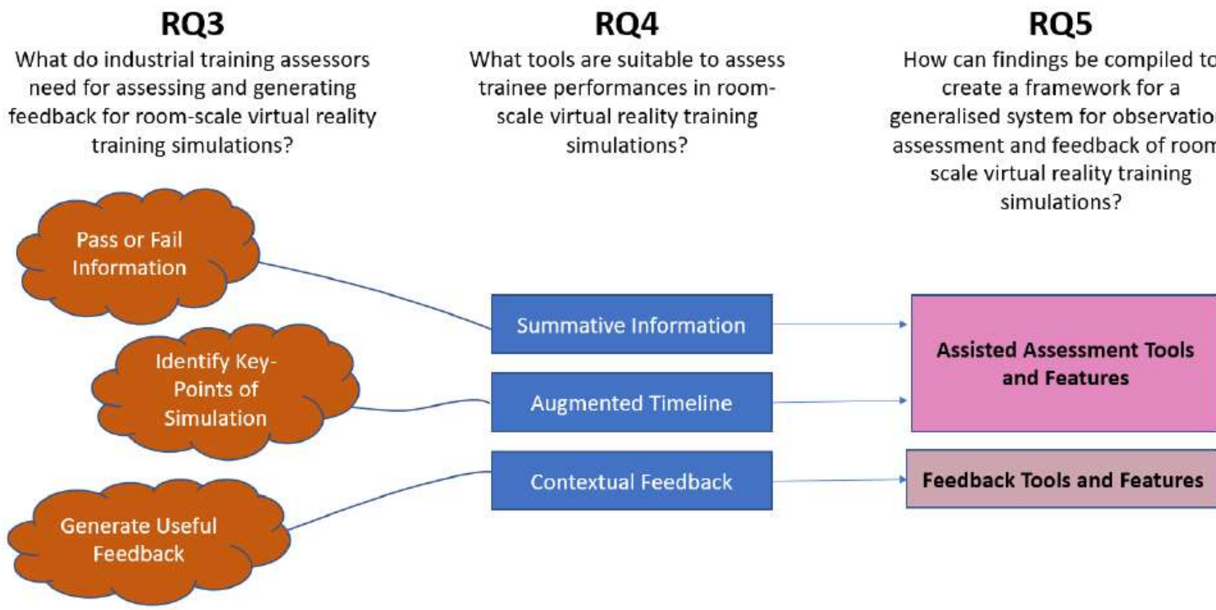


Figure 3.3: Evolution of the research questions from RQ3 to RQ5, demonstrating the process of analysing the interviews with assessors from the requirement collection study (see Section 4.1), identifcation of the requirements (see Section 4.2), and generation of the tools and features (see Section 4.3).

3.5.4 Research Question 6 and 7

The evaluation compared the framework against two existing methods discovered from the literature (RQ1), analytical data and video assessment. Six assessors from the oil and gas training industry and seven assessors from academia were recruited. During the study, assessors were observed using all three methods to conduct authentic assessment and generate contextual feedback for two virtual reality training performances. This was followed by semi-structured interviews, which allowed assessors to express their thoughts, feelings and impression of the three methods. The answers discovered for the following questions revealed the importance of elements within the methods which enabled a useful and usable method of conducting authentic assessment and generating contextual feedback.

RQ6: *Are the tools and features provided in the framework suitable for conducting authentic assessment and generating contextual feedback of room-scale virtual reality training simulations?*

Research question six evaluated the framework with assessors from industry and academia. The question sought to determine its usefulness and usability from the perspective of the assessor. This evaluation reviewed the components, features and tools, allowing understanding for how the framework met the needs of the assessors. Due to COVID19, the evaluation was conducted remotely using Microsoft Teams. Thirteen participants were recruited from industry (6) and academia (7) using the sampling approach discussed in Section 3.4. The evaluation of the framework was conducted to determine the following questions:

- Usefulness of the framework components, tools and features and interface.
- Usability of the framework components, tools and features interface.

The evaluation tasked assessors to conduct authentic assessment and provide contextual feedback for two VR trainee performances. After the evaluation, semi-structured interviews were conducted. Because the evaluation required open-ended responses, it was necessary to use a qualitative methodology to allow for in-depth responses to be generated from the evaluation. Because the evaluation wanted to not only determine how useful and usable the framework was for its purpose, but also determine what tools and features impacted the assessors ability to perform assessment and generate feedback and, why they did so. The semi-structured interviews enabled participants to recall from the evaluation and reflect on their experience to articulate their thoughts and feelings.

Thematic analysis was used to code and define the themes from the data. Once confident the themes defined the data, each theme was named to summarise the findings into their principal concept, with a thematic map generated of the final themes. These thematic maps produced the answer sought from the research question, focusing on the usefulness and usability of the framework and its components, tools, and features. Findings for the interviews are presented in Section 8.4.

RQ7: *How useful is the framework compared to state-of-the-art methods for conducting authentic assessment and generating contextual feedback of room-scale virtual reality training simulations?*

The final research question was conducted in parallel with RQ6, evaluating the framework against existing methods identified from the literature. However, unlike RQ6 which focused on the framework tools and features, this question sought to identify the strengths and weaknesses of each method for conducting authentic assessment and generating contextual feedback for room-scale virtual reality training simulations. This research question attempted to understand:

- Usability of the method for conducting authentic assessment of the trainee performance(s).
- Usability of the method for generating contextual feedback for the trainee performance(s).
- Usefulness of the method for conducting authentic assessment of the trainee performance(s).

- Usefulness of the method for generating contextual feedback for the trainee performance(s).

The participants, location, session and process for answering this question was identical to RQ6, as both questions were answered within the same evaluation session. The evaluation is presented in Chapter 8, with key findings in Section 8.3.

3.6 Requirements Collection Methodology

The concept of the novel framework steamed from an unpublished and unsuccessful first study the author of this thesis conducted to analyse remote virtual reality experiences for analysis of different movement methods for exploring an unfamiliar environment. The study created a virtual world environment for exploration and featured three methods of movement, linear movement, teleportation movement and a novel method which used a reference frame context to teleport the VR operator, relying on physical interaction with the movement method to function. The study was conducted in-person and online, allowing a wider audience to be reached with a larger sample size for analysis and results.

In-person, the study was captured to video following the same methodology discussed in Section 2.7.8. Observations were also logged by the lead investigator who was present during the study, taking note of interesting participant movements in real-world and within the virtual environment. The online portion of the study was hosted on the Steam Storefront (Valve, 2021) and advertised on VR community forms, being accessible to anyone who had a room-scale capable VR equipment. However, when the study was being transposed to operate online, many of the observational principles captured from the in-person study were not possible.

Because of this limitation, the study application was modified to automatically track and record the movement location every second for linear movement, or every teleportation for teleport forms of movement. This data was used to reconstruct the exploration of the participant, identifying how and where they explored the environment (see Figure 3.4). This approach was a mixture of Motion and Input tracking (see Section 2.7.7) recording and Environment Recording techniques (see Section 2.7.7), and was implemented to supplement the loss of the video capture from the in-person study. The tracking of data provided a structured posture of the participant, using the VR equipment to determine the head and hand positions (see Figure 3.5).

However, because the data captured from the online studies lacked sufficient depth, it was difficult to make conclusions with confidence from the data. While the data captured provided a reliable representation of the movement history of the participant, it lacked information the between movement data, including their perspective vision and reactions to the environment, both of which were easily identifiable for the in-person study. It become evident that framework might have been needed for conducting assessment for room-scale virtual reality simulations, this insight was confirmed by the literature review (see Section 2.7.10), which highlighted the limitations of existing

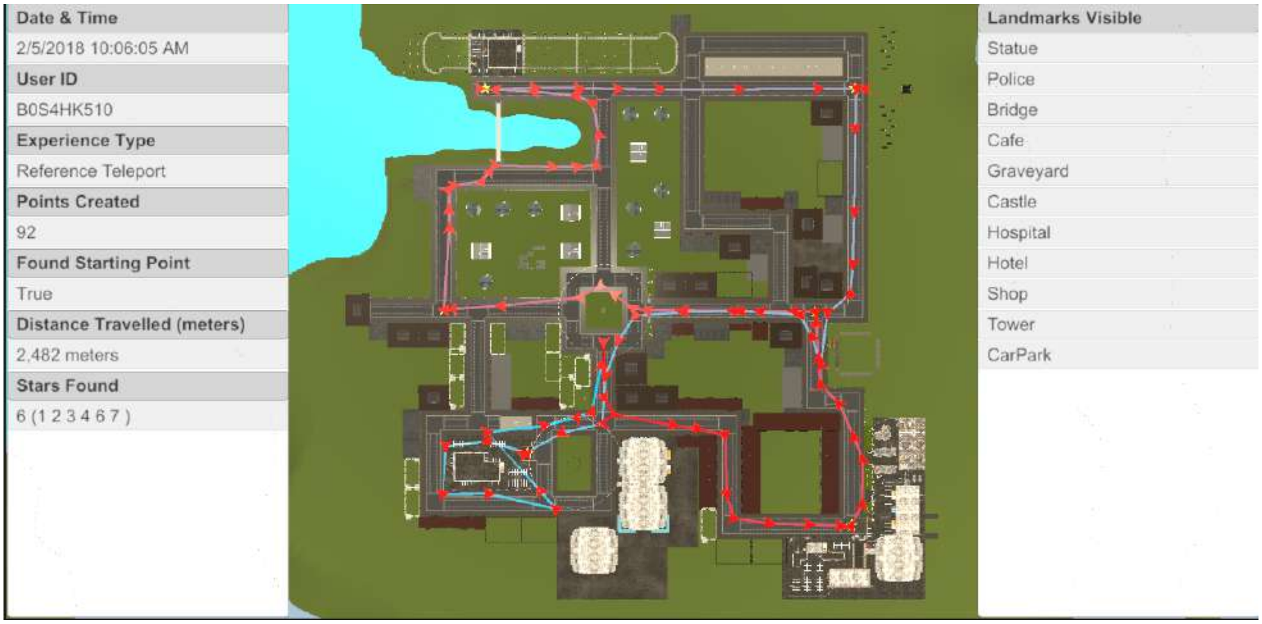


Figure 3.4: Reconstruction of the participant movement data captured from the online study overlaid onto the virtual simulation environment.

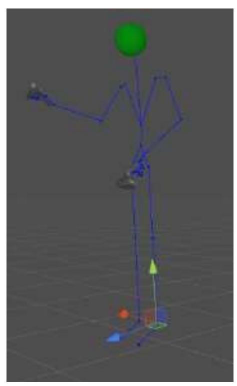


Figure 3.5: The capture of participant posture using the recorded location of the VR equipment. This implementation would evolve to eventually become the Skeleton Structure Visualisation tool within the framework (see Section 6.5.3).

approaches (see Section 2.7). As a consequence, this research focused on resolving this gap in knowledge.

While this study which was ultimately unsuccessful, it provided insight into the potential of VR studies, education and training for remote uses. Upon further research, it was identified that virtual training was the most immediately relevant subject area, therefore the focus was motivated on this area and a connection was formed with an industrial training company to identify the requirements of the framework within the field of VR training and assessment. This led to the framework focusing providing assessors with the capabilities of providing authentic assessment and generating contextual feedback for training simulations, replicating their approaches from classroom-based assessment.

Before attempting to create the framework, it was important to identify the needs and requirements from assessors, who conduct authentic assessment and contextual feedback trainees of in classroom-based assessment regularly. Semi-structured interviews with 5 industrial training assessors and field-observations monitoring classroom-based training and assessment at an industrial training facility was conducted. These participants were targeted for the study due to their knowledge and experience within the subject area, justified by the "information power" in Section 3.4.

A copy of the semi-instructed interview sheet can be found in Appendix D. Interviews lasted approximately one hour per participant, and were transcribed for analysis. Thematic analysis was conducted following the methodology presented by Braun and Clarke (2006), which revealed themes from the rich data, highlighting the importance of different elements of the themes which were extracted to form the characteristics of the framework component tools and features. This study is presented in Section 4.1.

3.7 Software Development Methodology

An agile development approach (Dingsøyr et al., 2012) was adopted to create the framework presented in this work to answer RQ5 (see Figure 3.6). This agile approach involved identifying the concept of the framework for authentic assessment and contextual feedback, monitoring technical feasibility (see Section 3.6) and incorporating findings from the literature review (see Chapter 2) and industrial training assessors (see Section 4.1). The framework concept was then iteratively prototyped to determined the technical viability of the approach for recording and reconstructing simulations, incorporating the requirements identified from the study with industrial training assessors for RQ3 and RQ4 (see Chapter 4).

After the technical evaluation of the framework concept proved viable, demonstrating that recording and reconstructing data was feasible (Howie and Gilardi, 2019; Howie and Gilardi, 2020), development work shifted to adapting the framework and implementing requirements discovered from the study with industrial training assessors (see Section 4.1). Because this was an iterative process, an agile approach was justified, with the framework evolving throughout the research.

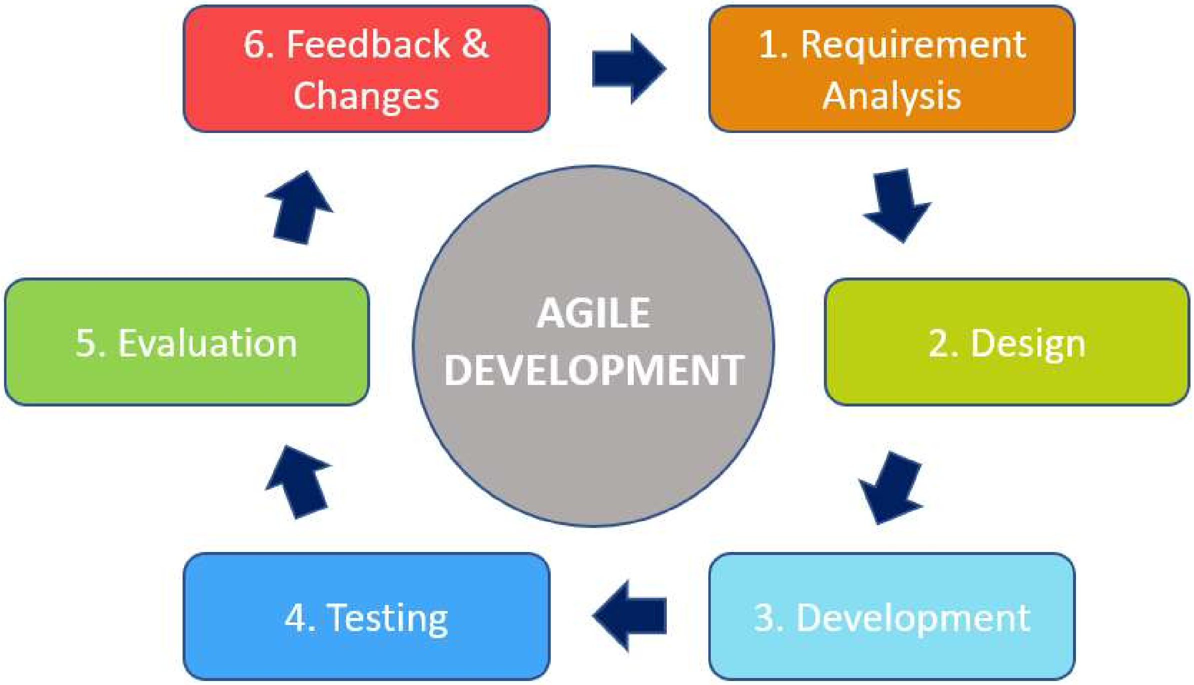


Figure 3.6: The agile development workflow used to design and develop the framework.

3.8 Evaluation Scenario and Process for RQ6 and RQ7

To evaluate the framework and answer the final two research questions of this work (RQ6 and RQ7), an evaluation of the framework and existing methods was conducted. To conduct the evaluation, a virtual reality training simulation was required to enable trainee performances to be captured and reproduced for assessment. The author developed a commercial oil-rig training simulator for an industrial training company, and through existing contacts with the company, permission was granted for the simulator to be used as the foundation of the training scenario (see Figure 3.7). This saved development time and allowed the assessment methods to be easily embedded into an existing authentic assessment and training simulation.

3.8.1 The Scenario

The *training scenario* used as the basis of the evaluation was designed to meet criteria determined by an industrial training company, which included 14 tasks that the trainee should complete in logical order. These tasks involve the trainee applying their skills and knowledge to overcoming the objectives, which included preparing the work area, identifying the fault, isolating the pressure line, dismantling the pipe and reporting the error to the in-game line manager. The tasks were:

1. Pick up the Radio: The trainee should pickup the radio to inform the in-game manager that they are ready to begin the simulation and to receive information regarding their objectives.

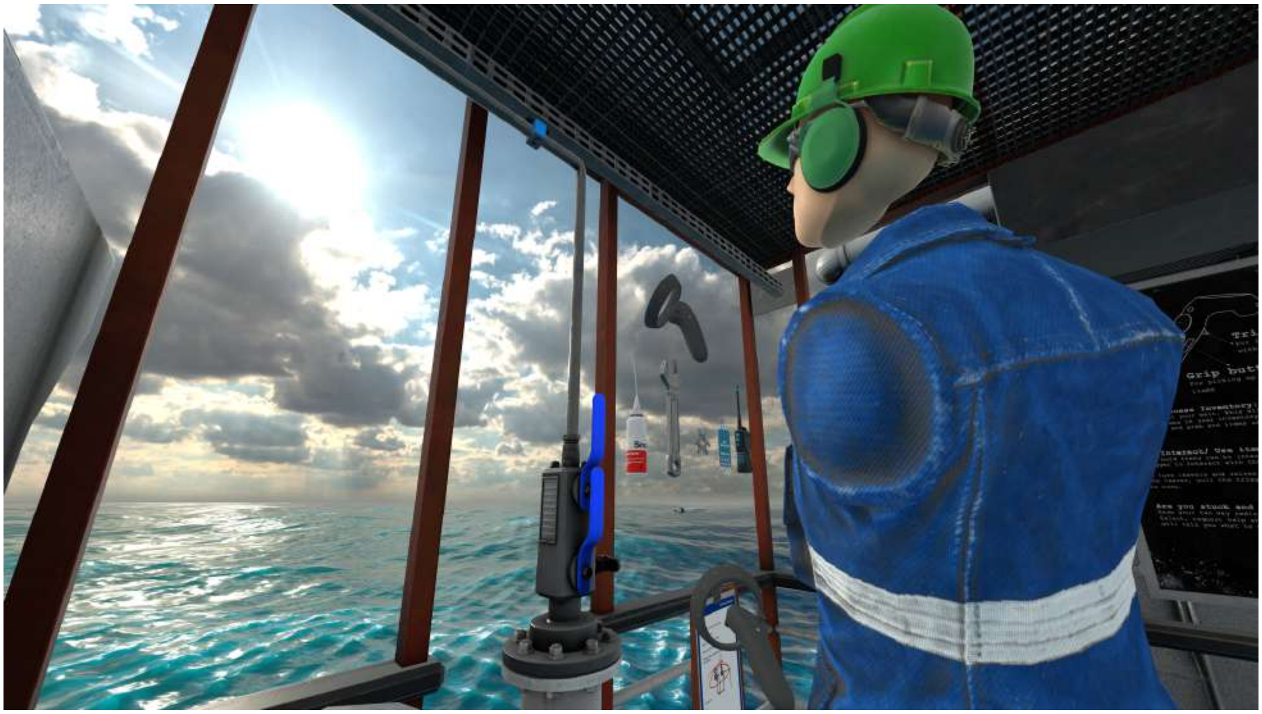


Figure 3.7: The oil-rig training simulation used as the simulation scenario for the evaluation (RQ6 and RQ7).

2. Pick up Clipboard: The trainee should investigate the fault detailed on the clipboard to identify the location and type of fault reported.

3. Clear Area of Hazards: The trainee should clear the work area of any hazards that makes the area unsafe to operate.

4. Barrier In Place: The trainee should place a barrier on the stairways to cordon off the work area before looking for any fault in the piping system.

5. Snoop Detector Hazard: Once the work area is cordoned off, the trainee should equip the snoop detector to identify the leak in the pipe located next to the vent gauge. When found using the snoop detector, the liquid applied on the area will display bubbles, indicating the leak and fault area.

6. Radio Report: The trainee should report the fault to the in-game manager using one of three available options. The trainee must respond with: "I've found the fault, I need to isolate PI1234 and investigate the leak. Am I ok to proceed?", choosing any other option is considered a minor failure, requiring the participant to re-try an alternative response.

7. Tag Fitted: The trainee should attach a Swagelok Intervention Tag to the work area indicating that the pipe system is under review and investigation by the trainee.

8. Isolation Valve Closed: The trainee should close both isolation valves that are connected to the fault area pipeline.

9. Depressing the Line: After both isolation valves are closed, the trainee should depressurise the line by operating either of the two venting valve gauges connected to the pipe system.

10. Checked Fitting Gap: Once the system has been depressurised, the trainee should investigate the gap of the nut using an inspection gauge. Doing so reveals the pipe must be disassembled.

11. Marking Pipe with pen: Prior to disassembly, the trainee should use the pen marker to mark the areas of the pipe that are being moved during disassembly.

12. Disassembled the fitting: The trainee should equip the spanner and use it to dismantle the pipe, which reveals the area of fault.

13. Identified Hazard: The trainee should identify the correct hazard from a selection of five and report it to the in-game manager. For the purposes of this experiment, fault "Debris on Ferrule" was defaulted for both data-sets, with the correct trainee response as: “There is debris on the front ferrule". Any other response should be considered a minor failure and the trainee is not allowed to change their response.

14. Identified Hazard 2: A second fault is present with in the work area, at the very end the in-game manager asks the trainee to inspect the work area to verify everything else is fine to re-open. Out of four available options, the fault "Heavy Pressure" gauge was selected for this evaluation. The trainee should identify the fault and report it to the in-game manager with the response: "There is a heavy pressure gauge on the end of a thin tube above the fault area". Any other response should be considered a minor failure and the trainee is not allowed to change their response. Afterwords, the simulation ends.

There are two failure conditions within the scenario, ‘Minor Error(s)’ and ‘Critical Safety Failure‘. A minor error is an incorrect action or response that is not considered a safety critical issues, i.e. the error is not dangerous, rather, it shows the trainee lacks understanding of the knowledge for the task. These minor errors should be corrected during feedback, but did not interfere with the training conditions. ‘Critical Safety Failure‘ errors are unsafe mistakes which will or may result in unsafe or dangerous consequences to the trainee, surrounding or operation of the oil-rig. If a ‘Critical Safety Error’ is triggered, the scenario immediately ends as the conditions of the scenario are no longer considered acceptable or resolvable by the trainee. See Figure 3.8 for more information regarding ‘Minor Error(s)’ and ‘Critical Safety Failure‘ errors.

Detailed guidelines were created for the purposes of evaluating the framework and were presented to each participant prior to the evaluation beginning, which functioned as an assessment sheet to guide the participants for the competency expected of trainee performance. The guideline

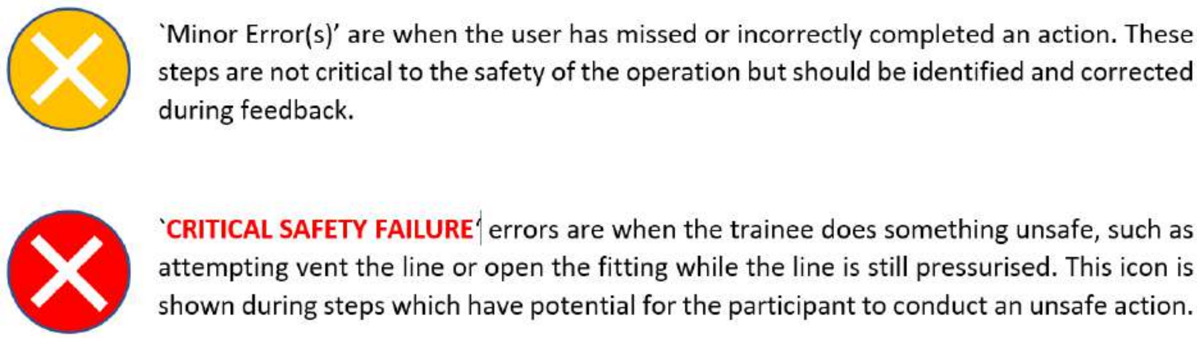


Figure 3.8: 'Minor Error(s)' and 'Critical Safety Failure' errors explained.

document was created to familiarise knowledge of each task and objective throughout the simulation in a style that enabled participants with no-prior knowledge of the simulation tasks and outcomes to understand and use them as a baseline for measuring trainee competency. By supplying the guidelines prior to the evaluation, each participant was able to identify the actions expected of the trainee at each task of the training simulation. The guidelines document is presented in Appendix C.

Two types of trainee performances were generated for the evaluation. These data-sets were fabricated to present areas of interests for evaluation purposes, with data-set A representing a successful scenario, while data-set B represented a scenario with a failure outcome. Data-set A includes a full completion of the simulation without any critical safety failure errors, but a few minor errors for steps missed to present opportunities for assessors (participants) to provide feedback to correct gaps in knowledge (see Figure 3.9). Verbal comments and visual animated cues (from the VR body tracking) were included in the data-sets to create opportunities for participants to identify any hesitation and to determine if they could acquire an impression of the trainee performance.

Data-set B involves a critical safety failure caused by the trainee not correctly isolating the both valves in Task 8 of the simulation, instead only closing one of the valves. When the trainee moves on to Task 9 to depressurise the line, the system explodes causing the simulation to end immediately (see Figure 3.10). Likewise, minor errors are included within the data-set (B) to provide other opportunities for the assessors to provide feedback. With regards to data-set B, two versions of the data-set were created, with the same conditions and outcomes, but with a modification of the timing of events and objectives completed. Data-set B-1 (DB1) was captured for the framework and analytical data, with data-set B-2 (DB2) for the video assessment method. Both conditions DB1 and DB2 share the same simulation outcomes as seen in Figure 3.10, with the only difference being the participant approach to these objectives. For example, DB1 was faster to identify the fault with the snoop detector, but DB2 hesitated for a shorter duration after identifying the fault and moving onto the next step. The two data-sets were created to identify if prior knowledge of the simulation data-set impacted their observation intention, knowing that data-set A was consistent, it

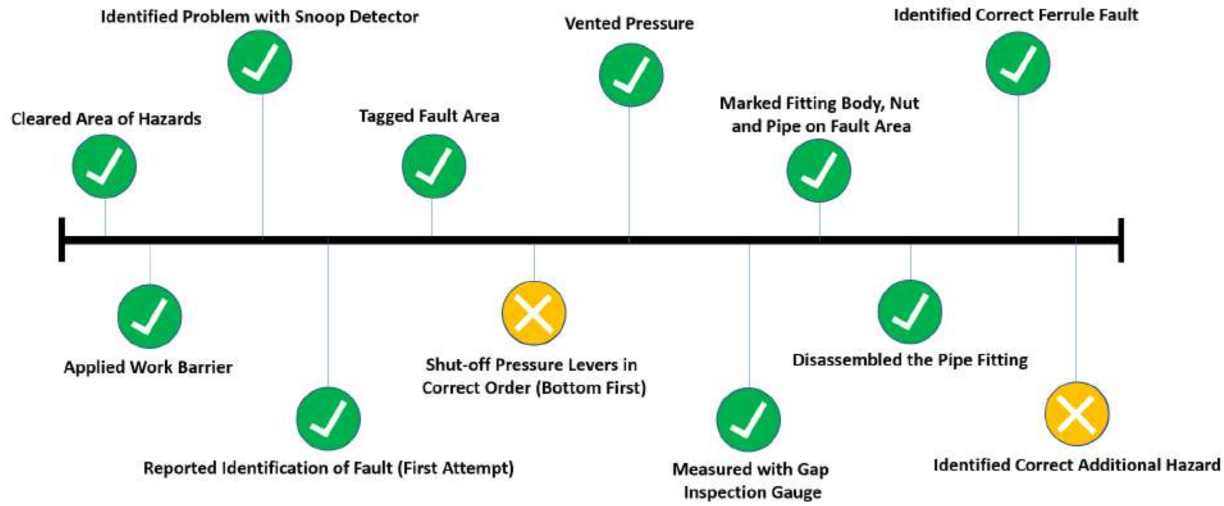


Figure 3.9: Data-set A conditions of trainee performance. The trainee A performance includes Minor Errors (yellow), but no Critical Safety Errors.

was anticipated the participant would expect the same for data-set B. Altering the condition enabled the second observation of the data-set B (1 or 2) to determine if the processes impacted how prior knowledge impacted use of the method.

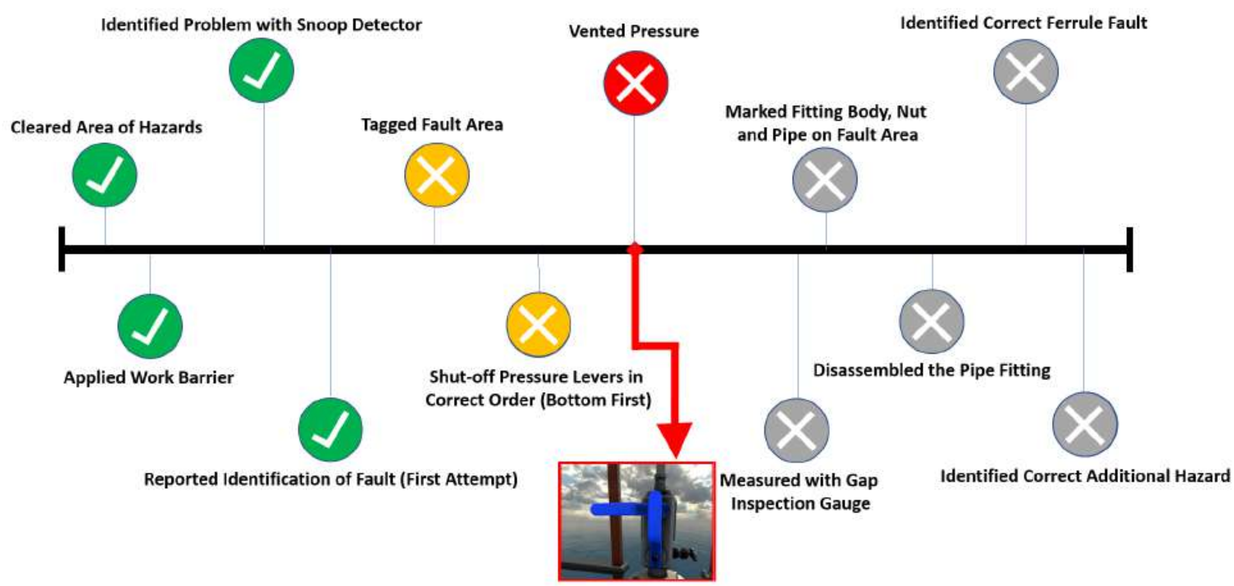


Figure 3.10: Data-set B conditions of trainee performance. The trainee B performance includes Minor Errors (yellow) and Critical Safety Errors (Red), with an overall failure outcome. Objectives not yet reached are Greyed out.

(a) Analytical Data-set A (b) Analytical Data-set B

Figure 3.11: (Left) Data-set A, (Right) Data-set B, as presented to the participants for the analytical method.

Existing Assessment Factors

Analytical data (see Section 2.7.2) and video assessment (see Section 2.7.8) were chosen as the *comparison methods* for the evaluation. These were chosen due to their accessibility for the evaluation. Other methods of assessment discussed in Section 2.7 were unavailable or unsuitable for use within the chosen simulation scenario. Analytical data was already embedded within the simulation as part of the training program, with its unmodified output forming the basis of the analytical data outcomes (see Figure 3.11). The video assessment method captured the trainee performance from their point-of-view (see Figure 1.1). Animation using Unity's game recorder was planned to be used as a comparison method during the evaluation, however when capturing all data from the simulation using the tool alone, with no other method operating, (see Section 2.7.6), the frame-rate of the simulation operated at 1 frame per second(s), and would eventually crash the software after a few minutes, capturing no data. For this reason it was discarded.

All three evaluation methods (the framework, analytical data and video assessment) captured data simultaneously when fabricating data-sets A and B for the evaluation, excluding the aforementioned differences for Data-set B. The video assessment method was captured using the Open Broadcaster Software at run-time of the simulation. Data-set A lasted 9:46 minutes, with a file size of 0.7 gigabytes (GB). Data-set B lasted (4:38) minutes with a file size of 0.43GB. Neither of these data-sets were modified or shortened length, therefore they include additional time prior to the simulation beginning and after the simulation ended until recording could be stopped. Complete recordings of the simulations captured using the framework were in total: 0.6GB for data-set A, and 0.22GB for data-set B.

3.8.2 The Evaluation Process

Participants were recruited via email and where assigned a time-slot of their convince between the 1st of November and 18th of December 2020, with the study being conducted remotely using Microsoft Teams. Each time-slot was booked on an individual basis, with each participant meeting to perform the study alone with only the lead investigator present. Informed consent from each participant was required prior to them conducting the study, with consent forms (see Appendix B) previously sent during the recruitment stage returned via email to the lead investigator. All participants volunteered to participate in the evaluation for no financial reward was offered. Factors were organised into a Latin square matrix (Lazar, Feng, and Hochheiser, 2017) (see Table 3.1). The order of participation automatically determined the order each participant would operate the methods, avoiding any potential bias by the lead investigator, and removing factor order bias from the participants.

Table 3.1: The Latin Square Matrix (Lazar, Feng, and Hochheiser, 2017) order during the study, which repeated for every fourth participant.

Participant Factor Use			
ID	First	Second	Third
1	The Framework	Video Assessment	Analytical Data
2	Analytical Data	The Framework	Video Assessment
3	Video Assessment	Analytical Data	The Framework

During the evaluation, the methods of evaluation were hosted on the lead investigators computer and were streamed to the participant over Microsoft Teams (MS) remote-screen sharing giving the remote participants full control over the interactions of the methods using MS teams control feature. However, TeamViewer was used as a backup for two participants who's hardware or operating system was incompatible with Microsoft Teams, functionality was nearly identical and this did not have any impact on the study. First, participants were asked to read the guidelines document to identify participant competency within the virtual training simulation (see Appendix C). After participants confident in their ability to conduct assessment of a trainee, the participant would move onto their first assigned method, first doing a five minute tutorial of the method if applicable. Operating the method, the participant would review two data-sets of trainee performance, followed immediately by one or more questionnaires about the method. A final questionnaire asked the participant to rank methods, then verbally explaining their reasoning of selection to the investigator after each question. A semi-structured interview with the participant concluded the evaluation.

The semi-structured interview focused on identifying opinions and feelings towards the different methods, along with attempting to identify the perceived advantages and disadvantages of each method and preference within the context of usefulness and usability, along with perceived problems. Questionnaires were used to develop the dialogue during the semi-structured interview, as the targeted sample size of population was insufficient for generating significant statistical results.

Three questionnaires were prepared for this evaluation. The first questionnaire, regarding usability was based off Lund's (2001) work and Brooke's (1996) System Usability Scale (SUS) questionnaire, which enabled a measurement of usability and usefulness of a product. The second was a usefulness questionnaires that was inspired by Chaudy's (2015) evaluation of the engage engine (Chaudy, 2015) and was used to determine participants feelings of usefulness of the components, tools, and features of the framework. The ranking questionnaire was used to directly compare the different methods of observation, allowing the assessor to select their first, second and third preference of method for each question. Interesting responses from the questionnaires were examined and revisited during the semi-structured interview to gather in-depth data on their reasoning and motivation behind responses. The semi-structured interview was conducted to allow participants to express themselves and generate in-depth motivation of their feelings regarding their participation in the study for evaluating the methods for authentic assessment and contextual feedback.

During the study audio was captured from the participants microphone, along with a screen-recording of the computer-screen using Microsoft Teams recording feature. Audio recordings of the entire participant session was automatically transcribed using Otter.ai, with the lead investigator correcting transcription mistakes immediately after the participant session and auto-transcription was complete. Notes taken by the lead investigator during the study was used to supplement and note areas of interest within the transcribed files during analysis.

Two areas of the trainee performance from both data-sets (one area for each) were selected as the points for the feedback to be generated by the trainee, to enable a consistent area of feedback to be measured between participants. These instances of feedback were initiated under predefined conditions: (data-set A) when the trainee expressed verbal (audio cues) and visual (movement cues) indications of the trainee failing to understand a task within the simulation, and (Data-set B) when the trainee conducted a safety critical failure. These instances were consistent in all data-sets (including DB1 and DB2). Participants were asked to review the entire trainee performance before to going back to provide feedback.

Due to time restrictions of the evaluation being around 90 minutes, participants were only asked to discuss their approach to providing feedback for the video assessment and analytical data methods. This was caused by the existing methods having no built-in method of providing feedback, requiring third-party alternative solutions. To avoid overwhelming and exhausting the participant, feedback for these methods were conducted by the participant describing the process of supplying the feedback along with its content step-by-step. When using the framework participants were asked to use any combination of embedded authentic assessment and contextual feedback tools and features to generate feedback.

Initial analysis of the data was conducted during the evaluation (via written observations) and immediately after transcription of a participant data-set to provide an overall summary of initial ideas and findings for that individual section of data. These initial findings were used to refine interview approaches to expand questions as part of the semi-structured interviews as initial theories and interesting results from the data emerged.

Data acquired from participants during the evaluation was analysed using thematic analysis, following the guidelines detailed by Braun and Clarke (2006). Detailed analysis of the evaluation was only conducted after participant interviews transcribed by reading the entire data-set for all participants twice, expanding upon initial ideas and notes generated during the first analysis of data, and generating initial codes within the data.

The coding process utilised a data-driven approach, with the research questions of this work used to form the relevance of how data is coded (Braun and Clarke, 2006), specifically regarding the usefulness and usability of the framework itself or as a comparison against other methods.

The coding of data was conducted using Nvivo 12, a qualitative data analysis software. All transcriptions of participants during the evaluation were imported into the software and manually coded using the built-in coding tools. Nodes within Nvivo were used to hold relevant coded sections which were then later incorporated into themes as part of the thematic analysis phases. The coding process was used to identify interesting and relevant features within the data that aided the researcher to understand perspectives of participants relative to the research question for determining the usefulness and usability of the framework itself and compared to the other existing methods (video assessment and analytical data). Codes were created to structure the findings from the data into subjects that best suited their relevance to the research question.

Once the coding of data was finished, themes were generated to merge coded sections of data. Generating themes was iterative, with new codes discovered and others deleted or merged, with the intention to ensure that data was not initially missed, ignored or miss-representative. Themes were regenerated, modified and deleted based on the changing structure of the coded data until the themes best represented findings for level 1 and level 2 of Braun and Clarke's (2006) phases for thematic analysis.

The generated themes were reviewed against the coded data to ensure they formed coherent patterns (Braun and Clarke, 2006), followed by the validity of the themes for how they present the data as a whole (Braun and Clarke, 2006). The identified themes were discussed with the lead supervisor and once confident the themes defined the data, each theme was named to summarise the findings into their principal concept, with a thematic map generated of the final themes. It was important to note that findings from the data were representative of the data, with themes having a coherent structure that formed the core of the theme.

3.8.3 Critical Discussion

The first potential criticism envisioned from the approach taken is the lack of input from the perspective of the trainee who operates the virtual reality training simulation. However, the focus and scope of this work is isolated on the development of a framework that meets the needs of assessors for conducting authentic assessment and generating contextual feedback for room-scale VR training simulations. Without first identifying the requirements of a framework for assessors, it is impossible to justify focusing on the trainee perspective, as the findings generated from this research are the necessary stepping blocks which allows for engagement with trainees in future work (see Section 9.3).

Secondly, the thematic analysis from the evaluation could have been validated further with the assistance of other researchers to form a group to jointly compare and analyse findings and themes from the data, instead of just discussing the results of the analysis with only the lead supervisor. However, while this would have improved the reliability of the data, it was not possible to have others involved in the research analysis stage due to the time constraints and the COVID19 pandemic. While the findings generated from this work are from the perspective of the author and lead supervisor, the process for analysing the data was rigorously adhered to, following the established principles by Braun and Clarke (2006), with findings and themes presented in this work representing the data discovered from the evaluation.

3.9 Ethical Considerations

Ethical approval was obtained for all experiments that involved human participants. Approved letters can be found in Appendix B. As this research required co-operation with human participants, ethics protocols were rigorously followed throughout the research to ensure that all participant information was kept confidential, secure and anonymous, following the ethical guidelines laid out by the University and ethics comity. All participants were over the age of 18 years old and did not demonstrate any mental or physical disability that would prevent them providing their explicit consent to participate, or to conduct tasks required during the study. During the Covid19 pandemic all studies were conducted remotely using a secure and private Microsoft Teams meetings to avoid health and safety risks for participants in face-to-face studies.

All participants were provided detailed information forms prior to their participation within the study, with each participant asked to read and digest the information carefully before proceeding. Informed consent via physical or digital signed consent sheets were required from each participant prior to their acceptance for conducting the study. Information and consent sheets are provided in Appendix B.

All recorded materials collected during the study were stored using private and secure cloud storage, with files password protected and encrypted to prevent illicit access and to keep materials confidential. Only researchers explicitly involved in the study and named within the ethics application were able to access this data. Participant information was kept confidential by anonymising collected data, detaching any identifiable information to participant numbers. Furthermore, all recorded material will be erased after five years from the submission date of this thesis to reduce risk of data breaches. While the deletion of all participant data after five years reduces the credibility and trustworthiness of this research in the future, it is important that ethics and good practices are followed to maintain a trustworthy relationship between researchers and participants.

3.10 Chapter Summary

The objective of this chapter was to outline the approach of the research methodology adopted to answer the research questions, including study procedure, participants and evaluation methods for qualitative thematic analysis.

This chapter has discussed the approach taken to conduct background research (RQ1) and determine existing methods suitability for conducting authentic assessment and contextual feedback (RQ2). Findings from these two questions (RQ1 and RQ2) identified that no suitable framework exists. As such, the methodology presented the approach used to conduct a requirement collection study with industrial training assessors to understand their perspective and requirements for authentic assessment and contextual feedback of room-scale VR training simulations (RQ3, RQ4). Following the identification of requirements, this chapter discussed the proactive methodology taken to design the framework (RQ5) and evaluate its usefulness and usability with assessors to provide authentic assessment and generate contextual feedback for trainees in room-scale VR simulations. The methodology presented the evaluation approach used to examine the suitability of the framework components, tools, and features (RQ6), and evaluate the framework against existing methods for conducting authentic assessment and generating contextual feedback (RQ7). The research adopts a qualitative methodology to answer the research questions, relying on semi-structured interviews and thematic analysis to generate findings from the data.

The methodology presented in this chapter underpins the approaches taken for this research project. Chapter 4 conducts a study with industrial training assessors to identify the requirements necessary to create the framework and its components, tools, and features, which is capable of allowing assessors to conduct authentic assessment and generate contextual feedback of room-scale VR training simulations.

Chapter 4

Framework Requirements

The following chapter presents the results from a study conducted with industrial training assessors from the oil and gas industry who use authentic assessment for classroom-based training and assessment. The study was conducted using semi-structured interviews and thematic analysis (see Section 4.1). Findings from the study were used to formulate a set of functional requirements (see Section 4.2) from which, the tool and features necessary for conducting authentic assessment and generating contextual feedback for room-scale VR training simulations were generated (see Section 4.3). This chapter answers RQ3 and RQ4 of this work: "*What do assessors need for assessing and generating feedback for room-scale virtual reality training simulations?*" (RQ3) and "*What tools are suitable to assess trainee performances in room-scale virtual reality training simulations?*" (RQ4).

4.1 Requirement Collection

As discussed in Section 3.6, a contextual study to identify the requirements of the framework was conducted with industrial training assessors using a qualitative methodology. Findings from the thematic analysis of five semi-structured interviews, supplemented with field-observations, identified three key themes: **observations**, **system features** and **concerns**. Results are reported and discussed below. The study was conducted using three principles: (a) identify the current approaches for classroom-based training authentic assessment and contextual feedback, (b) discover features or functionality that would be requested in within a authentic assessment and contextual feedback framework for VR simulations, and, (c) identify any concerns expressed by the participants in moving to a virtual medium for conducting authentic assessment and generating contextual feedback.

4.1.1 Observations

The *observation* theme (see Figure 4.1) identified several sub-themes which created overarching topics of discussion regarding trainee observations, assessments and procedures for conducting assessment. The observation theme focused on discovering how trainee performances could be reproduced in a format that is usable and useful for the assessors to directly observe and monitor trainee performances. The following sub-themes were identified under the observation theme:

1. *Confidence Level*: Determining the confidence level of the trainee assisted assessors identify trainees who were under or over-confident. Trainees who were under or over-confident were given additional attention during assessment and feedback.

2. *Difficulty Monitoring Trainee*: As trainees progressed through the assessment at different stages, tasks are sometimes obscured by the trainees body, making it difficult to keep track of the progress for each trainee. Assessors use a marking sheet during classroom-based assessment to keep track of trainee progress to identify the next set of actions expected of the trainee. However, in certain circumstances, it is possible that trainee actions can be missed or obscured.

3. *Task Repetition*: Assessors need to be able to identify quickly at what stage of the assessment the trainee is at. This knowledge is acquired from experience observing trainees conducting assessments.

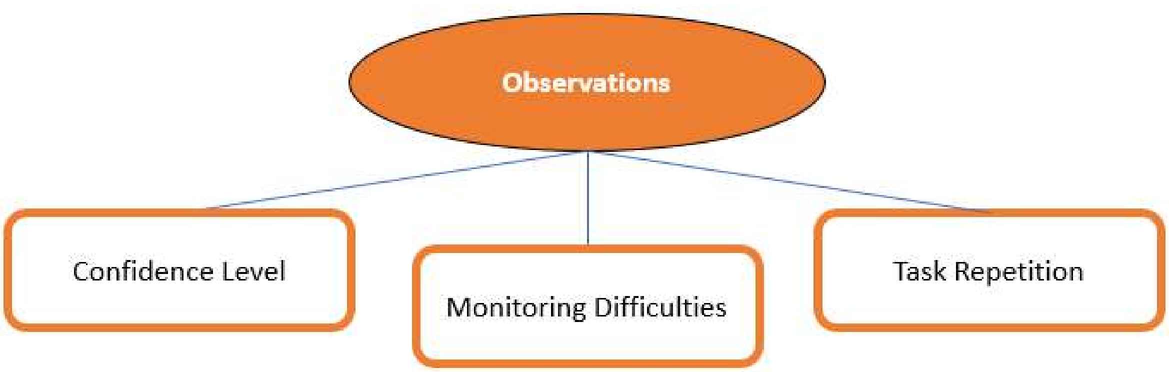


Figure 4.1: Theme and sub-themes for observation.

The themes elements revealed the process for how observation is currently handled during classroom-based assessment, providing information to the approaches and methodologies assessors adopt when monitoring trainees and how they judge trainee knowledge and competency. Classroom-based assessment used authentic assessment to measure the knowledge and competency of trainees during assessment, allowing trainees to demonstrate their skills and knowledge to complete a set objectives to overcome a problem (Wiggins, 1990).

One of the interesting findings from the observation theme was identifying how participants judged trainees, highlighting that the confidence level of the trainee played a key-role in determining how participants approached the observation. Trainees that showed signs of under or over-confidence were likely to struggle or make mistakes. As such, these trainees were given additional attention (monitoring) during the assessment.

Confidence was seen to play an important role in how assessors judged trainee's during assessments in classroom-based assessment, highlighting that confidence level played a key-role in determining how the trainee was anticipated to perform. Trainees who the participant felt were likely to struggle, were given additional attention during assessment and training sessions to monitor for expected mistakes. Confidence was measured by identifying the movement and actions of the trainee to determine if they are logical applications of knowledge:

> *"I would look to see that they have a confidence with what they're doing, so they seem clear in their direction of the activity that they're carrying out, and then you're observing against a set criteria so they're performing or achieving certain tasks or activities [...] there can be three levels of confidence. There can be people who are not confident, people who are confident, people who are overly confident and you can normally tell by how they fire into the practical." (Participant 4)*

Participant 4 discussed how they monitor the confidence of the trainee by focusing on the trainees approach to solving problems and comparing them to their expectation for an average trainee completing the task. Based on that criteria, the participant is then capable of judging if the trainee is confident, under-confident or over-confident:

> *"If somebody dives in that's an immediate this guy's just a bit cowboy, doesn't really know what he's doing as opposed to someone who's thinking things through first and you can see them having a look around, reading instructions properly, before they've even gone anywhere near the hardware. That's a big difference and usually if someone is doing that and you can see them thinking it through it's a good indication [...] How they approach it. I mean there's even a bit of body language in there. You're watching their body language," (Participant 3)*

While these confidence levels did not provide a definitive indication regarding a trainees pass or failure in the assessment, it provides assessors with an opportunity to identify a trainee that is lacking in knowledge, or allows assessors to speak to a trainee who is approaching the tasks too fast and not considering the health sand safety risks attached:

> *"I'm looking at have they got the understanding, the theory and the knowledge behind what they're doing, it's not just a robotic action that they're doing this. That when they're doing it they're checking things and they're understanding why they're checking these things" (Participant 1)*

Assessor look for trainee's that approach assessment objectives using a method that is safe and logical, using their time effectively to verify the work area is safe and operating normally. With correct understanding of knowledge, trainees are able to apply their skills to tackle an objective or problem correctly. As such, assessors require the ability to directly observe and monitor trainee apply their knowledge to determine trainee confidence and competency.

Monitoring Difficulties One of the potential, albeit infrequent problems with classroom-based assessment involved monitoring and observation of the trainee during the assessment. During the interviews, participants were asked to reflect on any past training or assessment session where they did or felt they may have missed an action, movement or identifier of a trainee performance. Participants reflected on the challenge of keeping track of the trainee progress, and suggest the loss of focus was caused by their attention being diverted by another trainees performance:

> *"You've been given more attention to somebody else for a particular reason then you might miss what somebody else is doing, I've definitely had that and then I've picked it up at the final stages but again we have our, like a last check to ensure that we can pick all things up." (Participant 4)*

One of the primary reasons expressed by Participant 4 for missing elements of the trainee performance is the challenge of monitoring multiple trainees who are all at different stages of their assessment. As an assessor, it is their responsibility to monitor each trainee, however, as up to four participants can be conducting the assessment at once, assessors attention is diverted between trainees. As such, while an assessor is monitoring one trainee completing an objective, another trainee may have conducted an error that initially goes unnoticed. In this case, the error would not be identified until the very end of the assessment when the practical work is dismantled and examined:

> *"As the guys go through and they start to progress at different stages it is very possible that you would miss certain little elements. There's things that you can catch a back up on, on the disassembly side but on the first initial assembly you may miss it and it's something then that you would not may be able to talk about at the end. You would just be assuming it was done right [...] You can't concentrate on four people all the time. If they're all doing different things you've only got one set of eyes and you can't really keep your eye on everything at the one time." (Participant 5)*

However, as noted by Participant 5, if something of interest is missed, such as a trainee attempting to use an incorrect tool, there is no way of identifying that if the initial observation was missed, preventing an opportunity for feedback that corrects trainees misunderstandings.

Because trainees can be at different stages during an assessment, to support keeping track of trainee tasks and objectives, paperwork is used to mark when a trainee has successfully completed tasks. However, even when using paperwork, it can be difficult to manage when monitoring four trainees at once and keep track of each individual trainees progress:

> *"you tick the boxes as they're supposed to complete those certain tasks and if they've completed those tasks they've passed, that's part of it. But if you're walking around with a sheet of paper and you've got five people you're not looking at 15 tick boxes and you're not looking at Bob, Bob just did item number four on this, tick. Oh Joe just did that one as well. You can't, you'd just be looking at a sheet of paper all the time and not watching them." (Participant 3)*

As Participant 3 suggests, with marking schemes the focus of the assessor is diverted further, requiring attention to shifted between the four trainees and four sets of marking schemes, causing further concern that trainee actions will be initially missed:

> *"I think the times you might have a challenge is if you're maybe having spent some time directly with one individual so maybe they've made a safety critical error and you have to go over and stop them but at the same time you've got three other people still continuing their assessments." (Participant 4)*

Participant 4 makes an interesting point regarding assessments where the trainee has completed a task that requires an instant intervention by the assessor related to a health and safety breach, again causing attention to be diverted away from other trainees conducting the assessment. Participant 4 continues:

> *"So, you might have to make a judgement call on should I stop that individual [who committed a health and safety error] because would that prompt the other individuals not to make the same mistake" (Participant 4)*

By intervening instantly when trainee has failed a task, the assessor has to use their judgement to determine if their intervention will prevent other trainees from making the same mistake, potentially impacting the other trainees assessment results. This suggests that it is best to conduct assessment as a one-on-one operation, however, that is not feasible in commercial classroom-based training due to the economic factors for facility, staff-costs and length of each assessment. One assessment observed during field-observations lasted the maximum available time of two hours, requiring an assessor to be present for a single trainee conducting the assessment.

Task Repetition When approaching assessments, assessors are able to efficiently conduct the assessment by analysing trainee performances using their experience and expectations from previously conducted assessments. The task repetition of the trainee performance provides a gauge for how well the trainee is coping:

> *"We do so many of the same assessments over and over again that you get a bit of a gauge with, even like a time point of view, so somebody, we would know what the average time would be to complete an assessment so you witnessed an assessment earlier today." (Participant 4)*

Participant 4 notes that the time a trainee takes to complete objectives during the assessment influences their expectations of the trainee performance, with past experiences providing reference:

> *"You know what to kind of look out for and you're almost waiting for them to make that mistake so you know where to sort of intervene or get involved before they make that mistake. You get to know, yeah, you always do that exercise. 70% of people always make a mistake at that so you're watching them at that point to try and catch it." (Participant 3)*

High probability of failure at certain areas of an assessment influences assessors observation during trainee assessments. When a trainee approaches an area of concern, assessors divert their attention to focus on that trainee. However, the challenge with classroom-based assessments is the presence of multiple trainees conducting the assessment at once, up to a limit of 4. As such, assessors are required to efficiently manage their attention between trainees, keeping track of each trainees assessment progress while monitoring the intricacies of trainee performances:

> *"You're standing back and taking in an overview, over watching everything that they're doing as a whole rather than focusing specifically on certain bits because again certain people might be at different stages so you can't just focus on that particular bit because they might be a stage ahead so you're trying to keep an over, it's having an understanding of the overall tasks, it doesn't matter what stage one guy is at." (Participant 3)*

Assessors note that it is important to keep an overall contextual understanding of the trainee(s) process of completing objectives for training. However, with each trainee assessment potentially lasting up to 2 hours, assessors workload is restricted, as an assessor must be present to observe and assess the trainee for the entire duration of the assessment. During these assessments there is a lot of down-time waiting for trainees to complete tasks, and conclude their assessment:

> *"There's a lot of waiting time where guys, well first of all when they're reading it they take a long time to study the task at hand and we tell them to make sure that they do that because don't just rush in, so make sure you understand what the task is. So, that might be 10 minutes if people stand there and read this piece of paper so you could just fast forward that." (Participant 5)*

Participant 5 expresses the desires to skip sections of the trainee assessment where the trainee is preparing, reading the instructions or familiarising themselves with the training environment. Because the assessor must be present for the full duration of the assessment, the process causes assessor to waste time from their schedule. This suggests that after-action review could be a more time-efficient method of conducting authentic assessment and contextual feedback, since sections of the trainee performance where the trainee is preparing or reading instructions, can be skipped, improving the time required to observe and conducted the assessment.

4.1.2 Framework Features

The *framework features* theme (see Figure 4.2) involved the tools and features assessors would require for a framework that would enable them to perform authentic assessment and generate contextual feedback in VR, which followed the same principles used in classroom-based assessment.

This theme composes of two sub-themes, which are composed of sub-sub-themes:

1. Assessment:
   (a) *Assisted Observation*: The ability to rewind the observation and view the trainee from any perspective, allowing assessors to identify and watch any and all objectives within a simulation. Furthermore, assisted tools and features should provide the assessor with ability to quickly identify the actions performed by the trainee
   (b) *Performance Summary*: Ability for the assessor to see an overview of the the trainee performance outcomes.
2. Feedback:
   (a) *Contextual*: The feedback supplied to the trainee must be contextual and meaningful for the trainee. Trainees must be able to understand where, when and why the feedback is provided, allowing trainees to learn and improve their skills and knowledge.
   (b) *Archiving Evidence*: Trainee performances should be archived in a format that enables assessors to analyse them in preparation of future assessments, or to provide evidence for situations were a trainee disputes the an assessment outcome.

The framework features theme findings identify the authentic assessment and contextual feedback requirements for the framework components, tools, and features. Each of the sub-sub-themes identified under the framework features theme is discussed the remainder of this sub-section.

Assisted Observation During classroom-based assessment, it was not always possible for the assessor to see all actions conducted by the trainee, with potential to miss areas of a trainee performance (see Section 4.1.1). Missing trainee actions was often the result of challenges monitoring and observing multiple trainees at once, with potential to lose track of individual trainees progress.

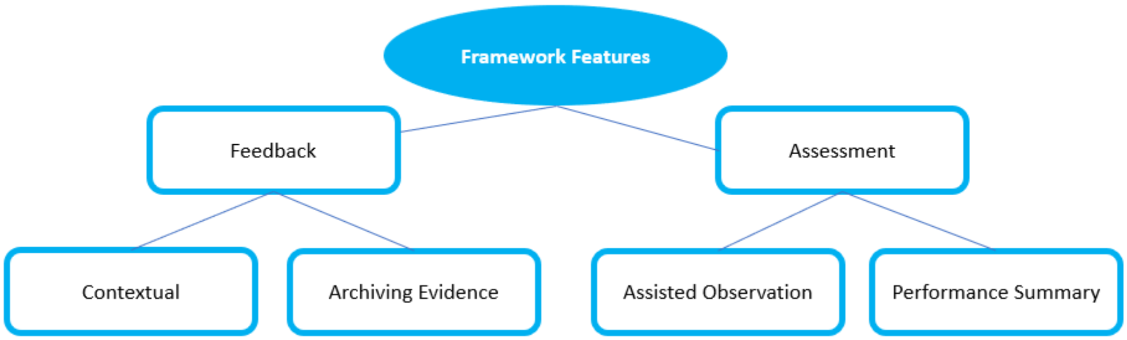


Figure 4.2: Theme, sub-themes and sub-sub-themes for framework features.

To gauge how assessment could be improved, participants suggested assistance features that aided their ability to conduct assessment and better observe the trainee performance. Participants referenced video as the ideal method of improving their observation and assessment performance, with multiple participants suggesting the inclusion of a replay system that enabled the assessor to re-check the video if they felt they had missed something from the trainees performance:

> *"If you were recording something I suppose you could play back certain elements of a, but I think at that point before you were playing something back somebody's maybe not understood something or they've done something wrong, so in that situation I would just show them it again." (Participant 3)*

Participant 3 suggests that the ability to playback video of the trainee session offers a better context to feedback, allowing the trainee to review their performance from video to identify what section of their performance the assessor is referencing when providing feedback feedback:

> *"You could rewind it, have a look, did that really happen or not. You would need multiple views which would be weird but I think that the good thing would be that you could potentially do it discreetly whereas if you were physically doing it and you need to have a close up look you might be bending over the guy or having a little stare at what they're trying to do whereas if you're using video you can zoom in, zoom out." (Participant 5)*

Participant 5 explains that with the inclusion of multiple video views, the observation problems of assessment would be reduced as the assessment can be remotely conducted to view all perspectives of the assessment without interfering the trainee, allowing the video to zoom in on points of interest.

During classroom-based assessment, assessors position themselves in an area of the room to observe all trainees at once, without intruding on the trainees work-space. Because the trainees assessment outcomes are often requirements for jobs and certifications, most trainees have some anxiety or nervousness:

> *"I suppose you would be able to have a different viewpoint so you could potentially have the view of the person who's carrying out the activity I suppose. I'm sure you can move around 360 degrees and all angles zoom in and out. I think for me that's something because as an assessor the last thing I want to do is be standing over somebody's shoulder when they're trying to carry something out." (Participant 4)*

> *"It's difficult without being intrusive as well because they feel, it's a bit like a driving test or something, they're nervous." (Participant 5)*

Participant 4 and 5 both express a concern with current observation, as they do not want to observe a trainee too closely as it could cause unnecessary pressure on the trainee, causing the trainee to panic and make a mistake. By using a remote form of observation, from a multi-view video perspective, the trainee would be unaware of when and where assessor watching.

These results suggest that the observation and assessment of a trainee can be improved by allowing alternative options for viewing the trainee performance, which enables verification of trainee actions and prevents elements of the trainee performance from being obscured. From the perspective of industrial training assessors, it can be concluded that video forms of observation and assessment appear viable solutions, which would improve the assessment process for both remote and classroom-based observations of trainees.

Performance Summary As discussed in Section 4.1.1, assessors were found to focus on specific elements of a trainee performance relative to their task on during the assessment. When monitoring a trainee the assessor is relying on their expectations and experience of the assessment processes to determine what stage the trainee is at, but when monitoring multiple trainees at once, it is easy to lose track of each individual trainees assessment progress:

> *"If you had four guys there and they were all doing the same activities step by step at the same time, right, that would be looking at four guys doing the same activity and ensuring that rig was safe. If you've got to look at four people at the same time. That's a difficult impact. That's what I'm trying to convey here." (Participant 2)*

Participant 2 notes that because all of the trainees conduct the assessment at different speeds based on their individual skill-levels, it is easy to lose track of what stage and task each trainee is on. Assessors reflected on how they would opt to improve the activity monitoring of trainee actions. The common response was to categorise the trainees performance into sections for observation, with each section focused on one task or objective of the trainee performance, allowing quick and easy identification and validation of what task the trainee is doing, what their progress is and what their next action should be as part of their assessment:

> *"I think if you see somebody who is reading the sheet all you need to say is they've read it. You don't need to watch them reading it. They might take 10 minutes to read the instructions. I don't need to watch them reading it for 10 minutes. All I need to do is say tick they've read it and then if I can skip through to the next action and observe that, so I think if the program was able to identify an action and it would be that there was a trigger in there that says that's the next action then I think that would be safe." (Participant 4)*

Participant 4 discusses with an identifier that represents what the trainee is doing at sections of their performance, the assessor would skip sections that did not need assessed. In the example provided, Participant 4 reflects on an earlier response from Section 4.1.1 to highlight that they would use an action identification system to determine when a trainee is conducting assessment tasks and when they are reading instructions. With this information the assessor would skip to only the relevant parts of the assessment where the trainee is conducting actions that are part of the marking scheme. Participant 4, then further expands upon the idea of an action identification system, demonstrating its potential use to spotting irregularities in a trainee performance:

> *"I think if they picked up a spanner and let's say they undid something, and then they picked up another tool, even if they didn't use it I think you would still want to see that happen because you want to say well why are they picking up this other tool, what are they trying to do or thinking about doing [...]. Did they try to do something and then change their mind and that's something there, because that might prompt the assessor to want to ask that candidate some questions to understand why they did certain things." (Participant 4)*

In the example Participant 4 provides, they explain how the use of tools and other irregular activity of the trainee would be of interest to them for assessment purposes, as it provides a quick insight into the overall actions conducted by the trainee. In this example, the action identification system would then require an ability to monitor a wide variety of factors within the trainee performance to determine and identify what actions the trainee performed. However, Participant 4 concludes by observing that while an action identification system could be used to identify and categorise different sections of the trainee performance, it is not a replacement for directly observing the trainee:

> *"I don't think you would be able to skip the whole thing and just watch very tiny little bites, you would still have to look generally just to make sure that their approach is correct." (Participant 4)*

As such, a performance indication system system should be used to supplement the assessment process, providing assessors easy and quick insight into a trainees overall performance and actions performed.

Contextual Feedback The ability to provide feedback for the trainee performance is crucial to the learning experience for the trainee. In classroom-based feedback, assessors would use physical gestures and other visual cues to focus trainee attention. However, when asked to consider how they would adopt their feedback to trainees who conducted their assessment remotely, participants stressed the importance of supplementing the feedback with sufficient information to replace the use of traditional in-person gestures:

> *"So, I suppose in that scenario you're into the realms of having to be super clear with what words you use because you're not in the room to use hand gestures and direct and point, so you're having to be super super clear with the language you're using and if you're in the room you can probably skip certain words and use gestures of some description. But if you're just using [a simulation], you're not in the room, you're purely reliant on clear and concise verbal instruction." (Participant 3)*

Participant 3 stresses the importance that feedback supplied remotely is clear and concise, defining exactly what the target of the feedback is and providing a contextual presentation to the trainee. The feedback must be able to stand on its own, with clear information that is understandable and useful for the trainee to digest and learn from. When referencing sections of the trainee performance for feedback, the assessor should make full effort to contextualise the feedback in a manner that is understandable and meaningful to the trainee. However, given the complexity of training simulations, it may be too difficult to rely on verbal feedback alone. As such, it may be necessary to supplement the feedback other elements, such as video of images of the objectives or equipment.

To counter-act the lack of physical presence when providing feedback, verbal or written feedback should incorporate contextual elements that enable the assessor to clearly and concisely contextualise the feedback to targeted areas from within the trainee performance:

> *"I mean some people have got different learning styles haven't they so they take on things differently. Some folk they just prefer to see something, to be able to touch something, a combination of both, so I'm sure something visual would definitely help." (Participant 5)*

Because trainees learn differently, it is important to provide as much context to the feedback as possible to make up for the loss of any physical demonstrations or gestures that may be used in classroom-based feedback. When generating feedback, it is important that feedback is supplied with visual context providing the opportunity for the assessor to provide targeted feedback that focuses on the identified concerns from a trainee performance.

One participant suggested using the video timeline from a trainee performance which was recorded to video as a means of providing feedback. The timeline timestamp would then provide a visual reference to contextualise the feedback:

> *"You would probably have certain timelines in said video where you put in comments at this stage you did this, you know, at three minutes 37 seconds you turned this valve handle in the wrong direction when you should have blah blah. Something like that." (Participant 3)*

The suggestion appears to complement early proposals of providing context to the feedback. By connecting feedback to target the exact instance in a trainee simulation, it enables the assessor to clearly and concisely describe the fault relative to the visual cues, while the trainee is able to re-witness their actions, receiving relevant feedback to support their leaning. Building upon this idea, Participant 3 goes further, suggesting the use of video to provide the feedback in a single element:

> *"[For feedback] you could overlay [voice]. I suppose you could watch the video and overlay voice over it after the event or you could type annotations or something [into] the video." (Participant 3)*

Participant 3 conclude their suggestion of feedback by highlighting that assessors could overlay the feedback directly onto the video, and beyond that, annotate sections of the video to provide a relevant and contextual feedback source that can be easily transmitted to a trainee remotely. While feedback is provided to support the trainee learn from their mistakes, it also provides the opportunity for the trainee to reflect on their training performance. As such, reviewing videos of trainee performances would enable both assessors and trainees to walk-through the trainee performance together:

> *"He had connections that were leaking because he hadn't tightened them correctly so it was quite clear to him the error that he'd made but he didn't understand why he'd made that error. Now, I could see from observation why he'd made it so I knew before I even test it that it was probably going to leak but I wanted to carry out that test so he could see it leak as well. So, he could understand the consequence of his action [...] We physically showed it on the actual system what had happened and the reason it was loose, why that had happened and how he could prevent that from happening." (Participant 4)*

Participant 4 notes, he as an assessor was able to identify the point in the assessment when the trainee failed, however it was unclear to the trainee. Rather than stopping the assessment, the participant let the trainee continue until the failure caused a leak in the system. With this approach the trainee was able to see how his actions and input during the assessment impacted the outcome of events. After the error had revealed itself, the assessor was able to step in and explain how, why and where the mistake happened which caused the trainee to fail the assessment. By providing contextual and targeted feedback overlaid onto a recorded trainee performance, the trainee would be able to use the feedback for self-reflection.

Overall, these results for feedback highlight that feedback for remote trainees must be contextual and references the fault or area being discussed. Secondly, verbal feedback conducted for remote trainees should contain some visual element to clarify the focus of the feedback, ensuring the trainee is able to understand it and use it to improve their overall knowledge. Finally, video feedback appears a viable option for providing assessment to trainees remotely. However, when providing feedback using video, the assessor should be able to narrate, annotate or overlay information onto the video to contextualise the content of the feedback. The feedback should enable the trainee to reflect on their experience and provide a source of learning.

Archiving Evidence During assessments, the outcome of the assessor is final and is judged based on the checklist that confirms the trainee completed each objective of the assessment correctly. However, participants spoke of occasions when trainees refused to accept a failure outcome, and would escalate the case challenging the assessors judgement. As such, assessors discussed the use of capturing and archiving photographic evidence of a trainee failure:

> *"The best way is definitely a factual way. So, something hard and fast with evidence is the best way because when there's something without evidence it's subjective and that's the issue, and that's where you get challenged on it. So, something that's really factual and you've got evidence to prove it seems to take away any of those issues." (Participant 5)*

As a precaution, assessors would capture photographic evidence of the failure to archive it incase the trainee followed up on their threats for disputing the outcome. As noted by Participant 5, capturing the trainee mistake creates evidence that can be factually proven for any follow-up dispute that may arise, enabling the assessor to present the trainee failure, supporting the assessors original judgement:

> *"So, we've got a copy of the observation records, depending on the situation we've had, so when I've had it and that's absolutely happened what I've ended up doing is photographing the practical and it's stored photographs of their completed work with the paperwork so that we've got a reference to go back to." (Participant 4)*

Participant 4 expands upon the discussion, indicating that a copy of all trainee assessment outcomes are kept to enable an archival database of past performances. This archived data can either be used to counter disputes, or for assessors to familiarise themselves with a trainee who re-attempting an assessment.

4.1.3 Concerns

The final theme discovered from the study was regarding the *concerns* of training and assessment within VR (see Figure 4.3), with emphasis on authentic assessment and contextual feedback.

Concerns are an important consideration when designing the components tools and features, as it is important that the framework provides a level of confidence for the assessor, rather than increasing doubt in the feasibility or viability of virtual reality as a training medium. It is also important to attempt to minimise these concerns by adapting the framework design to provide assessors with a trusting and reliable method for conducting authentic assessment and generating contextual feedback for VR training simulations which replicates classroom-based assessment approaches. The following sub-themes of concerns were identified:

1. Computer Feedback Unsatisfactory: Assessors expressed concerns regarding the feasibility and reliability of the assessment conducted and feedback generated by AI or automated systems.

2. Abuse/Cheating: Assessors were worried that VR assessment would be prone to cheating, with trainees abusing bugs, glitches or limitations in the simulation that reduce the would limit capabilities of assessors conducting authentic assessment of the trainee.

3. No Instant Communication: When conducting after-action assessment, rather than real-time assessment that assessors use in classroom-based assessment, assessors were concerned that they would be unable to instantly intervene or provide feedback for trainee performance.

4. Passive Observer: Assessors were concerned that there role as assessors would be reduced to being passive observers, with no ability to provide useful input for assessment or suitably support the trainee with meaningful feedback.

5. Hesitant of Technology: Assessors expressed concerns with the viability of VR technology for trainees to perform authentic assessment that is similar to classroom-based assessment due to lack of tactile feedback and sense of weight when operating equipment in VR.

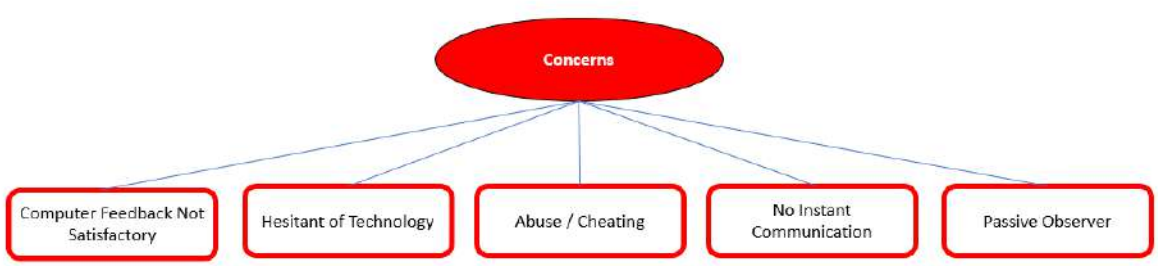


Figure 4.3: Theme and sub-themes for concerns.

It is important to discuss these sub-themes to understand how the framework component tools and features can be designed to best mitigate assessors concerns.

Computer Feedback Not Satisfactory Participants did not trust the assessment outcomes automatically generated from VR training simulations. Using knowledge from their own experience using VR training simulations, participants were concerned about the lack of reliability from the predefined outcomes:

> *"Again, it would essentially be an observation. The program itself can obviously record if a task has been completed or not completed. I think the limitations of the software means that somebody, there's only so many things that somebody can do unless you build all these into the system. So, you know, unless you make every single valve openable or closable that you can see there's limitations there. So, that's probably a bit of a challenge but I would say that the program would be able to record results but I think you would still have to have somebody who's physically overseeing that candidate go through the motions and the steps. So, whether it's live observation or after the fact, so they walk through the play a bit afterwards but I think somebody would still have to observe it." (Participant 4)*

Participant 4 highlights that for simulations providing authentic assessment, there are several potential avenues the trainee could utilise when tackling an objective. This is important as the assessment should provide the ability for direct observation of the trainee so that assessors can visually see the application of knowledge. With direct observation, assessors are able to identify why the trainee failed, rather than just being told where the trainee failed. As such, feedback supplied from direct observation can be explanatory. However, predefined outcomes can still have a use within VR training, as it provides immediate estimation of the trainee performance outcomes:

> *"Yeah, pending, you could put the areas that are pending and just say that these require observation by an Examiner or something like that, so it's not a you've failed but it's a you're going to have a discussion now with an Examiner and we're going to go through it and you'll get a result. " (Participant 4)*

As Participant 4 notes, automatic assessment would still require visual examination and verification by an assessor through direct observation of the trainee performance. Automatic and predefined outcomes should only be used as an outline of the anticipated performance of the trainee, which an assessor can use to familiarise themselves with the trainee performance prior to direct observation. This mixed approach provides the assessor with instant information for where the trainee failed in a simulation, with direct observation identifying why and how, allowing assessors to provide targeted contextual feedback.

Abuse/Cheating Concerns were expressed for virtual reality assessment, with Participant 2 suggesting that the format would be too open to abuse and enable a trainee to cheat their results:

> *"In practicality that be like saying to the kids when they're doing their O Levels or whatever they do these days "right you can sit alone and fire in your test paper" you know. Somewhere along the line there's going to be abuse for that system." (Participant 2)*

Participant 2 compared the remote virtual training with the learning outcomes from student exams if they were conducted from home and out of sight of an assessor. The participant expressed that he did not think remote students could be trusted, and if given the opportunity, would abuse the remote examination to cheat the marking schemes. In the example given, Participant 2 was reflecting on learning outcomes that were automatically determined by a virtual training system, without any assessor involvement in the assessment. To mitigate this concern, trainee performances within VR training simulations must be directly observable by an assessor, allowing validation and verification of trainees logical application of knowledge to completing objectives. With direct observation of the trainee performance, trainees should be unable to abuse automatic assessment marking.

No Instant Feedback Instant feedback was considered an important aspect of classroom-based training, with participants discussing the importance of providing feedback to the trainee after their assessment, settling then and there the pass or fail outcome. Switching to virtual remote training however would require a delay between the trainee conducting the assessment, and the assessor generating the feedback. As a consequence, it was suggested that this time-gap should adopt a temporary feedback evaluation result that was pending an assessor overview of the training outcomes:

> *"I don't know if it would be possible for the system to understand if you put it on clockwise, anti-clockwise, or your actions. I think it's good that they get the instant feedback, the decision is made there and then. They don't walk away and then you get told two days later that you didn't pass it and here's the script why." (Participant 4)*

While instant feedback would not be a comparable replacement of an assessor, the feedback can provide an instant outcome that is pending validation by an assessor using direct observation of the recorded trainee performance:

> *"I think the risk of doing these offline observations is the candidate will be given instant feedback but actually that feedback could be subject to change after somebody has reviewed it." (Participant 4)*

However, this approach may mean the instant feedback is not reliable, as an assessor may judge the performance differently, spotting mistakes an automated instant feedback system did not. As such, this approach may cause trainees to initially think they have passed, and only later discover their assessment attempt was a failure. Consequently, if instant feedback is provided to the trainee, it should be subject to an estimation only, with the final outcome pending validation by an assessor.

Passive Observer Another concern expressed by participant was the lack of engagement with the trainee when using remote forms of assessment and feedback. In classroom-based assessments, the assessment is conducted by the trainee with an assessor present in the same environment observing the entire process. As such, the assessment outcome and feedback are provided as soon as the trainee finishes their assessment. Feedback is then followed by a short question and answer session if the trainee wishes, providing them with the opportunity to discuss their performance with the assessor and ask any relevant questions or seek support. However, when asked to consider how assessors would conduct authentic assessment and generate contextual feedback in after-action situations when the trainee is not present, assessors were primarily concerned about the lack of engagement possible:

> *"You can't ask them questions. So, I think you need that ability to be able to say to the candidate why have you decided to do it this way, or what was the reason you picked up that tool. These kind of questions to ascertain their knowledge and understanding." (Participant 4)*

Switching from classroom assessment to remote assessment conducted using video caused assessors to suggest they would become passive observers in the assessment process, rather than proactive assessors. Participant 4, expresses that using a video to conduct assessment feels more abstract, since they as assessors will be relying on less information from trainee cues to determine the trainee performance. Likewise, the lack of these cues makes it difficult to generate feedback, as they are unable to discuss the performance with the trainee to identify significant mistakes. Participant 4 suggests that this could be resolved using a live video call. However, as the focus of this research follows Hanoun and Nahavandi's (2018) after-action guidelines, emphasis was placed on training, assessment and feedback situations that are non-instant. Potential for live streaming trainee assessments and follow-up dialogue sessions are discussed in Sections 6.5.4 and 8.3.4 respectively, mitigating Participant 4's concern.

Hesitance of Technology Suitability While most participants welcomed the adoption of technology, one participant did express a concern about moving training and assessment to a virtual medium. The reluctance for feedback stemmed from a hesitance of the virtual technology, caused by a concern that them as an assessor would be unable to adopt it:

> *"The technology is be far beyond me, I mean I was at university and we were the first class to ever get email." (Participant 2)*

While these claims are valid, Participant 2 expressed during the interview that they were not confident in the technology as they did not understand it, therefore, they did not fully trust it.

As such, it will be important the adoption of VR into teaching, training and education is made accessible to assessors from any technical background. This can be achieved by listening to these concerns and seeing what can be done to resolve or mitigate them. One potential solution is to conduct further studies with assessors using VR to acquire rich qualitative data that provide insight from their perspective, allowing researchers and developers determine how their concerns could be overcome. While these are interesting results that may be relevant to future research in this direction, the results are beyond the scope of this research aim.

Another concern expressed was the viability of virtual reality as a training medium, since the virtual format was unable to replicate all assessment elements from classroom-based training:

> *"They can do optic feedback but you're not going to have that sense of weight, so it's going to be sort of simulated [...] I mentioned that before where you physically pick something up so you're still picking things up, something that has a weight to it. So, you could pick up a bar that's simulating the torque wrench." (Participant 4)*

Again, participant 4 makes an interesting point, highlighting that virtual simulations struggle to simulate all aspects of realism training. The actions and motions of a torque wrench simulated with present technology do not replicate its true sense of weight and tension. There are a few potential avenues for overcoming these concerns. One could be to investigate the effects of substitutional reality training Simeone, Velloso, and Gellersen (2015) and virtual reality training to determine if the learning outcomes and performances of trainees are significantly different. If so, further research could focus on resolving that gap in knowledge. Alternatively, the use of haptic feedback and physical VR equipment interaction, like hand tracking or gloves, could provide a better replication of the senses for using the equipment. These changes in simulation interaction may prove sufficient to tricking the trainee through immersive and physical cues that create the illusion of physical interaction (Jerald, 2016). However, as this research focuses on a framework for providing authentic assessment of VR training simulations, the work to address the concerns expressed by Participant 2 and 4 is outside of the scope of this research. These concerns were included in the research to demonstrate that virtual training still requires further work regarding for its adoption within industry training.

4.2 Functional Requirements

The themes presented in Sections 4.1.1 and 4.1.2 can be used to define the requirements for tools that enable authentic assessment in room-scale VR, ensuring that such tools and features satisfy the needs of assessors to: (a) adapt classroom-based assessment to VR training and assessment, (b) conduct authentic assessment, and (c) generate contextual feedback.

The aim of authentic assessment is to observe and judge the trainee against a set of criteria to determine the trainee competency (Wood et al., 2013). To achieve this, tools will have to:

R1 provide direct and unrestricted control of the point of view in the trainee performance without hindering the trainees ability to operate the training simulation and complete the tasks.

R2 enable the assessor to make such judgement by monitoring and observing the trainee applying their knowledge to the tasked objectives.

By providing assisted assessment aids that streamline the assessment for the assessor, the capabilities and opportunities of the assessor should be enhanced, through improved observation and assessment capabilities. The assessor should be able to:

R3 quickly identify the task the system logged as passed and failures.

R4 identify key-points of the trainee performance to speed up or skip down-time in the simulation when the trainee is reading instructions and yet to start any objectives.

R5 identify the confidence of the trainee, so assessors can provide additional supportive feedback for trainees that appear to be under or over-confident.

Feedback enables trainees to reflect on their experience with the input from an assessor used to improve trainee confidence, and correct mistakes or gaps in knowledge. In classroom-based training both the assessor and trainee would be present within the same context, however with virtual feedback being conducted as after-action and detached, trainee and assessors would be in different contexts with the risk that the nuances of physical gestures from real-world feedback would be lost. To mitigate this risk, feedback must be able to:

R6 generate contextual evidence for the trainee that presents a clear indication of the feedback for when, where and why it is being generated.

R7 a record of trainee performances for be kept to be used either as evidence for disputed results or to monitor trainees who are reattempting assessments.

R8 supplement for the lack of gestures enabling the assessor to clearly and concisely focus the attention of the trainee on exact feedback points.

These functional requirements answer RQ3, presenting the tools and features that are considered necessary within a framework for assessors performing observation, conducting authentic assessment and and generating contextual feedback.

The next section will answer RQ4 by generating the tools and features that define the details of the framework design from Figure 1.6.

4.3 Generating the Tools and Features of the Framework

To answer RQ4, *"What tools are suitable to assess trainee performances in room-scale virtual reality training simulations?"*, the requirement collection study with assessors identified the functional requirements of the tools and features required within the framework (see Section 4.2). These functional requirements are used to generate the tools and features that are necessary within a framework to enable assessors to conduct authentic assessment and generate feedback for room-scale VR training simulations.

These tools and features are split into three components, providing tools for: *observation*, *assessment* and *feedback*.

4.3.1 Observation Tools

The observation component previoulsy shown in Figure 1.6 is intended to convert the recorded data into a visual and interactive format which reproduces the trainee simulation in format that is usable for assessors to directly observe and monitor the performance of a trainee in room-scale VR. Assessors must be able to directly observe, monitor and assess trainee's virtual performances to measure authentic assessment(Wood et al., 2013; Wiggins, 1990).

From the requirements discussed in Section 4.2, there are two tools required to conduct observation of a trainee performance in VR: a *free camera tool* and *replay features*.

Free Camera Tool The first requirement (R1, see Section 4.2) was the need for direct and unrestricted observation of the trainee performance. To enable assessor to observe the trainee performance, a free camera tool should be embedded within the observation component of the framework. The free camera tool should enable observation of the training simulation from any perspective, influenced by the free camera tool utilised in existing implementations of virtual performance observation (Stone, Snell, and Cooke, 2016; Raij and Lok, 2008). The tool should enable the assessor to have complete and unrestricted control of their observation, allowing them to track the trainee in room-scale movement and ensure that their vision is not obscured or restricted, as was found to be a minor problem in classroom-based assessments.

Replay Features The second requirement (R2, see Section 4.2) stressed the need for assessors to directly monitor and observe a trainee apply their knowledge to the tasked objectives. Like many existing implementations of after-action review methods (Stone, Snell, and Cooke, 2016; Raij and Lok, 2008; Thurston and Martin, 2011), the observation component should include the ability to pause, rewind and skip time(s) or trainee actions during the observation. This feature should enable the assessor to observe the trainee at any time, from any perspective. To achieve this, the framework should capture all data from the trainee performance to enable non-linear reconstruction, allowing trainee performances to be consistently reproduced. With replay features, the assessor should be able to review and re-watch areas of the trainee performance in-case something was initially overlooked or obstructed, which was a limitation and concern for classroom-based assessment.

4.3.2 Assisted Assessment Tools

Assisted assessment tools overlay additional features and elements onto the direct observation of the trainee performance, providing the assessor with additional information and insights into the performance of the trainee. An example being trainee search patterns (Stone, Snell, and Cooke, 2016) or timestamped transcription of participant audio (Raij and Lok, 2008). These tools and features are designed to assist the assessor, making their job easier for conducting assessment (Hanoun and Nahavandi, 2018).

From the results identified in Section 4.2, three tools and features were identified as requirements to conduct assessment: *Summative Performance Indicators*, *Augmented Timeline*, and *Assisted Visualisation of the Trainee Tools*.

Summative Performance Indicators The third requirement (R3, see Section 4.2) identified the importance of trainee performance outcomes for objectives which are logged as pass or fail. Access to these summative performance indicators can enable assessors to familiarise themselves with the trainee performance, providing prior knowledge of the expected trainee outcomes of the performance being assessed. Unlike traditional classroom-based assessment which is synchronous, assessment and feedback in VR is conducted after-action (see Section 2.7.1). As such, assessors can familiarise with the trainee performance outcomes through summative performance indicators that present the assessor with an overview of the trainee performance. Summative performance indicators should allow the assessor to focus additional attention during assessment to areas of a trainee performance where a failure indicator is shown by the summative performance indicator. This allows the assessment to improve upon the limitations imposed from real-time assessments conducted during classroom-based assessments.

Not to be confused with automatic assessment, the summative performance indicators should identify from the trainee performance in the simulation, the tasks and objectives which were registered as completed successfully, or failed.

Augmented Timeline The fourth requirement (R4, see Section 4.2) showed the potential of identify the key-points of data within the simulation, allowing assessors to minimise down-time by skipping or speeding up areas of the performance where the trainee is static, such as reading instructions.

To achieve this, an augmented timeline should be used to present the key-points of information of the trainee performance against the time of data modification, creating an easy to digest method of identifying key-points during the trainee performance. Raij and Lok's (2008) work demonstrated an effective timeline of trainee dialogue, allowing the assessor to jump to timestamped comments for trainee relative to the simulation environment. As such, the augmented timeline should provide the assessor with an an easy to view modification of key-points from trainee data, such as actions conducted by the trainee, which is presented in sync with the direct observation features (see Section 4.3.1). The augmented timeline allows assessors to quick identify and validate any area or key-point from a trainee performance, improving the assessment capabilities of assessors compared to classroom-based assessment.

Assisted Visualisation of the Trainee Tools The fifth requirement (R5, see Section 4.2) was the ability for assessors to identify the confidence of the trainee. With posture, movement and verbal cues from the trainee providing an estimation of confidence (see Section 4.1.1), it is important that these cues are represented in VR, providing real-world context to the actions of the trainee in the VR simulation, as was presented in (Howie and Gilardi, 2019; Howie and Gilardi, 2020). To achieve this, the assessor should have tools that visualise the trainee posture and movement, providing a simple skeleton structure visualisation of the trainees stance and movement while completing actions. This should be paired with the audio captured from the trainee VR equipment microphone, allowing assessors to hear and see what the trainee is doing throughout the simulation. These assisted visualisation of the trainee tools, should enable assessors to acquire comparable information about the trainee that would be expected in classroom-based assessment.

4.3.3 Feedback Generation Tools

Upon completion of assessment, the assessor should generate feedback that supports the learning of the trainee (Wijewickrema et al., 2017). Feedback is given to correct trainee misunderstandings, with the hope the trainee will reflect and learn from the feedback, improving their learning outcomes and maximising the learning benefit from the VR simulation (Raymond and Usherwood, 2013). As such, tools are required that enable assessors to generate contextual feedback that supports the trainee.

Findings from Section 4.2 provide an indication of the tools and features assessors would expect when generating contextual feedback in VR. The three tools generated from the requirements were: *Video Narration Tool*, *Evidence Capture Tool* and *Drawing Tool*.

Video Narration Tool The sixth requirement (R6, see Section 4.2) showed the need for feedback to be contextual, so the contents of feedback were clear and comprehensive, detailing when, where and why the feedback was generated. To provide contextual feedback for VR simulations, video was considered as the best method of generating and distributing the feedback to the trainee. As discussed in Section 2.8.2 video is a referenced and targeted method of providing feedback, which includes the context necessary to identify when and where feedback was generated within the simulation. For why, the feedback requires the assessors comments to be embedded. To achieve this, a video narration tool should be designed which enables assessors in real-time to record areas of the trainee performance (to video), capturing the context of feedback, with assessors providing the content of the feedback through audio narration. The resulting output of the video narration tool should be a single standalone video file, which includes a video capture of the trainee performance with assessors audio feedback embedded. This means trainees should have the visual and verbal elements of the feedback to comprehend when, where and why the feedback was generated. With these contexts, the video narration tool should replicate how feedback is currently supplied during in-person classroom-based assessments. The video narration tool should be compatible with all other tools and features.

Evidence Capture Tool The seventh requirement (R7, see Section 4.2) identified the need for archiving evidence. In Section 4.1 assessors demonstrated instances with trainees who disputed their assessment outcomes, requiring evidence to be collected. To replicate classroom-based approaches of storing evidence, an evidence capture tool should be provided, enabling assessors to capture an image of the the trainee within the simulation environment, clearly displaying any disputed errors made during assessment.

Drawing Tool The eighth requirement (R8, see Section 4.2) presented the need for a tool that can be used to supplement the gestures used when generating feedback in classroom-based assessment. In Section 4.1.2, assessors highlighted that gestures are used to focus the attention of trainees during feedback. This is in line with the literature in education (Valenzeno, Alibali, and Klatzky, 2003). As it is difficult to replicate physical gestures and cues virtually, a drawing tool is required to supplement requirement R8. The drawing tool should offer the same capabilities of gesture cues, allowing assessors to focus the attention of the trainee similar to classroom-based assessment feedback. The content drawn from the the drawing tool should be functional with all other feedback generation tools *video narration tool* and *evidence capture tool*.

4.4 Chapter Summary

This chapter answered RQ3 and RQ4, identifying the needs of industrial training assessors for conducting authentic assessment and generating contextual feedback for room-scale VR training simulations. Using the requirements identified from assessors (see Section 4.2), the tools necessary for observing the trainee, conducting assessment and generating contextual feedback were generated (see Section 4.3).

These tools and features will be designed and embedded into a framework that seamlessly merges together the requirements and tools necessary for conducting authentic assessment and contextual feedback in room-scale VR training simulations. As such, the tools and features generated in Section 4.3 should be compatible and support each other, allowing an overlapping of requirements that maximises the tools and features impact for enabling observation, assessment and feedback.

The first step in creating the framework was the design of the recording components necessary to capture data from trainee performance in VR simulation that would meet the requirements identified for RQ3 and RQ4. As existing methods were inadequate (see Section 2.9), a new approach was required. To solve this, the framework created a dual-recording methodology to capture all data from a VR simulation. Chapter 5 discusses the design of the dual-recording methodology and the first two components of the framework, lightweight recording and processed recording.

Chapter 5

Dual Recording Methodology

5.1 Introduction

To meet requirements identified from assessors in Sections 4.2 and 4.3, the framework required the ability capture data for trainee performance within VR simulations. However, as identified within the literature (see Sections 2.7.6, 2.7.7 and 2.9), current approaches to recording have several limiting factors which impact the run-time simulation performance or do not capture sufficient data (Backlund et al., 2007; Lopez et al., 2017) that provide the functional requirements needed by assessors from RQ4 (see Section 4.2) to provide authentic assessment. For instance, linear data recording (Von Spiczak et al., 2007) is based on motion capture techniques that record only the trainee data, but not the additional environmental data that is required for authentic assessment (Backlund et al., 2007; Lopez et al., 2017). Another example is the impact of recording on the frame-rate (Oculus, 2021) when using reflection techniques to record or reconstruct data, as discussed in (Howie and Gilardi, 2020).

The difference between the framework and existing approaches such as Howie and Gilardi (2020) lies in the approach to capturing data. Not to be confused with video recording implementations (Thurston and Martin, 2011), the framework is based on environment recordings (see Section 2.7.7) and operates by recording the modification of virtual data embedded into a simulation environment, similar to work by Raij and Lok (2008), Stone, Snell, and Cooke (2016) and Backlund et al. (2007).

This chapter discusses the patent-pending dual-recording methodology (Howie and Gilardi, 2021) that it is used to capture all the data in the simulation without impacting the performance of the simulation at run-time. The data captured by this dual-recording methodology enables the tools and features for observation, assessment and feedback to meet the requirements (see Section 4.2) identified from the requirement collection study with industrial training assessors (see Section 4.1).

The framework features two components for capturing data: lightweight recording and processed recording. Lightweight recording captures user input and non-deterministic data, while processed recording captures all remaining deterministic data within the simulation. Figure 5.1 presents an expanded view of the framework previously presented in Section 1.6, highlighting the role of the dual-recording methodology discussed in this chapter for the lightweight and processed recording components.

As over 50% of VR applications are developed using the Unity game engine (Latham et al., 2019), the design and implementation of the framework and its components, tools, and features targeted the Unity game engine. As such, the design of the components will focus on C-sharp terminology, as C-sharp is the primary programming language when developing with the Unity game engine.

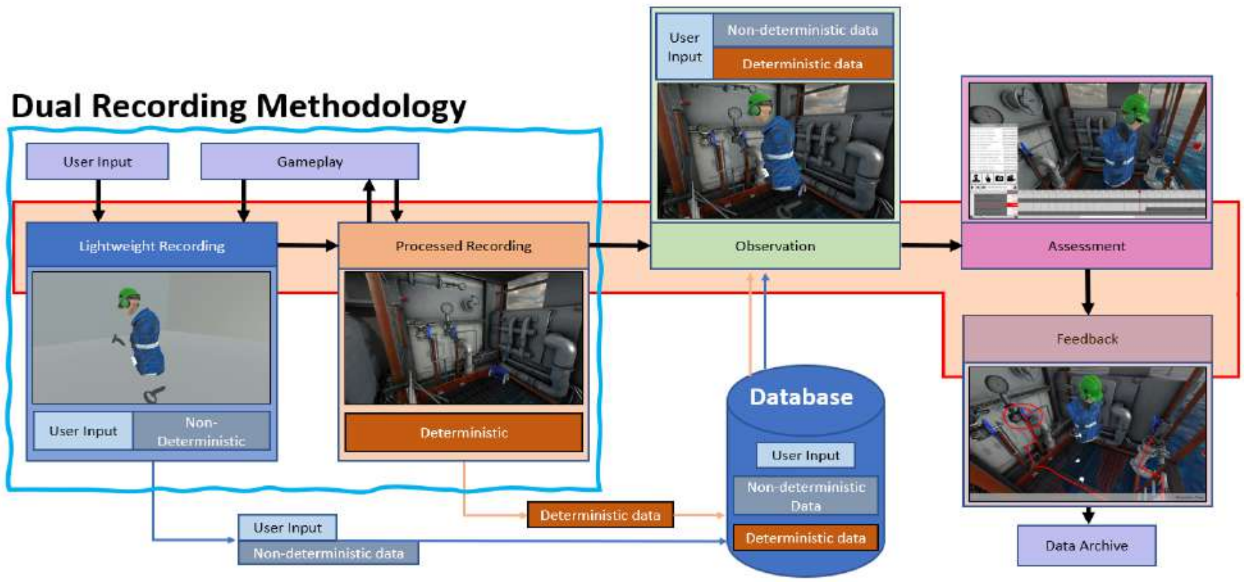


Figure 5.1: An overview of the framework showing the dual-recording methodology that is the focus of this chapter. This shows an expanded view of the framework overview first presented in Figure 1.6.

5.2 Framework User Interface

Figure 5.2 presents the user interface (UI) of the framework embedded within Swagelok Scotland's oil-rig VR training simulation. This user interface provides the display and navigation features to assessors who are using the framework to observe the trainee, perform authentic assessment and generate contextual feedback. The UI which includes the timeline, summative performance indicators, tool options and playback features is overlaid in-front of the observation free camera tool, providing in-context information of the trainee performance which is synced to the time of observation. The interface was designed to replicate the interface found in video editing suites, keeping all track (displayed items) and timeline information at the bottom of the content being observed.

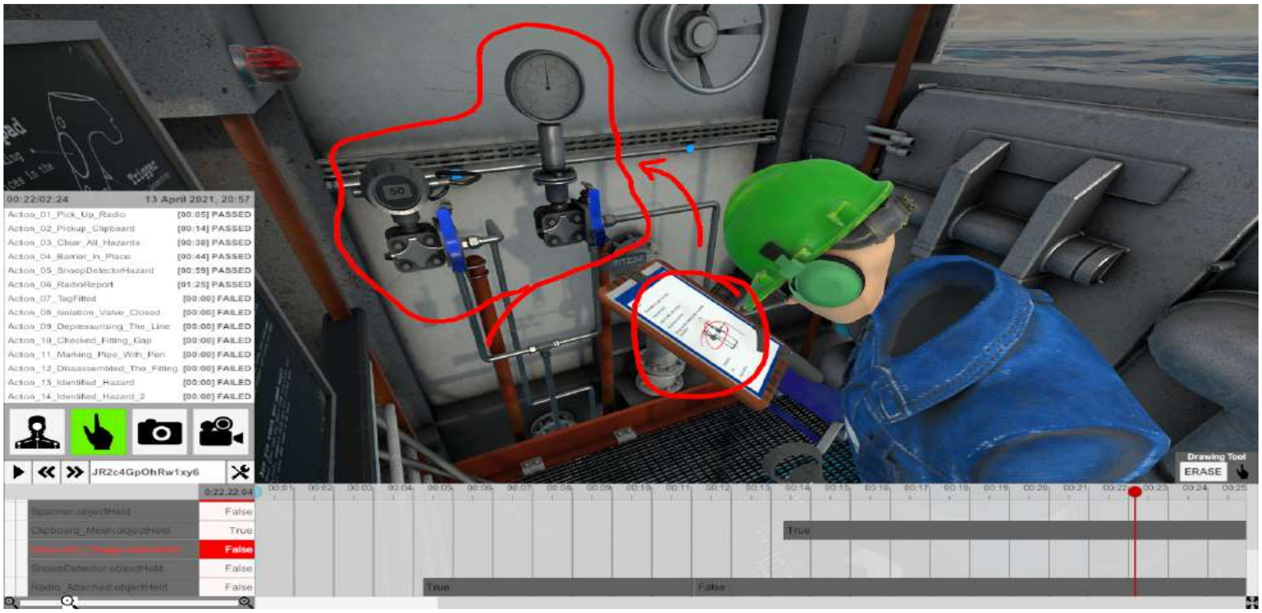


Figure 5.2: The end-user interface of the framework embedded into Swagelok Scotland's oil-rig VR training simulation, showing assisted assessment tools and features, including the augmented timeline, summative performance indicator and drawing tool.

Figures 5.3 and 5.4 present a documentation view of the framework interface, which labels each UI element and correlates to a tool or feature that is presented in the design of the framework. The playback tools include the ability to play/pause, rewind and fast-forward, along with functionality to alter the observation time by dragging the red keyframe marker to a new point on the timeline (see Figure 5.4). The UI for the framework also includes functionality to alter the content displayed on the timeline, using the horizontal scrollbar to display variable value changes throughout the timeline of the simulation. The vertical scrollbar is used to scroll through the list of equipment that is displayed on the timeline, allowing the timeline to scale to suit the needs of the simulation.

Details regarding the content of the timeline interface is presented in Figure 5.4, showing the equipment from the simulation embedded into each timeline row with status blocks which represent the value of the variable. In this example (see Figure fig:timelineobjectsmarked), the equipment variable displayed on the timeline is a boolean for determining if the equipment is currently held by the trainee. However, the timeline status blocks can be representative of any value that can be visualised using a string variable type, such as int, float, bool, enum, or vector, just to name a few examples.

As the framework is designed to be applicable for any VR simulation, the contents of the framework are dynamically created to scale and adapt to the content being displayed, meaning the same layout of the UI elements will remain, but the items presented in the equipment (object and variable) rows will adapt to alternative content and variable types and values. The process for designing and developing the framework to provide the tools and features that are presented on the UI will be the focus of the following two chapters, signified by the dual-recording methodology presented in

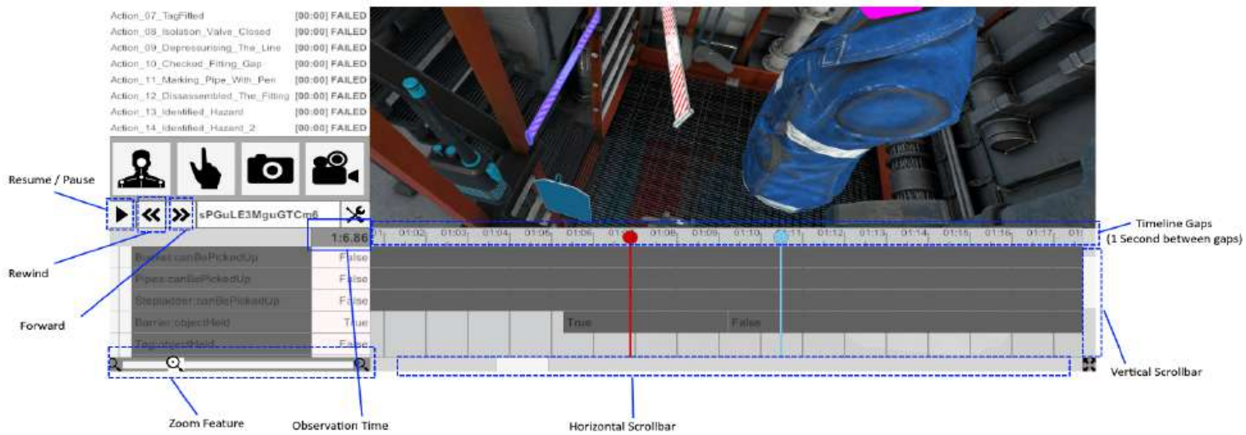


Figure 5.3: The UI of the framework showing the playback tools and user-interactivity of the interface for interacting with the timeline.

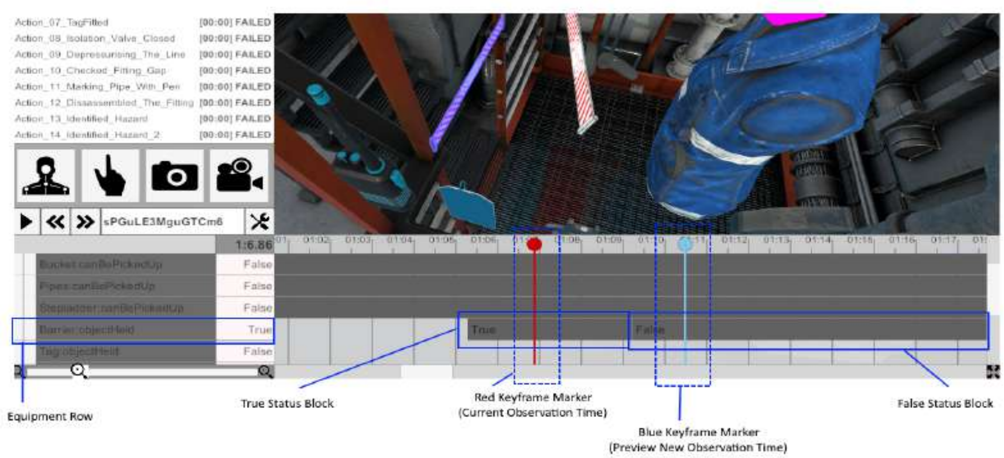


Figure 5.4: The UI of the framework showing timeline elements which display equipment (objects from the simulation), a selected variable value, and the status of that value throughout the simulation. In the example shown, the boolean value (true/false) is shown for the equipment barrier, with variable indicator objectHeld, which identifies if the object is currently held by the trainee.

this chapter, and the design of the remaining framework components, tools and features in Chapter 6.

5.2.1 Oil-Rig Training Simulation Scenario

The framework was designed to support any simulation, game or application that uses the Unity game engine. While this work has focused only on the oil-rig VR simulation developed by Swagelok Scotland, the framework has been incorporated into a fire training simulation in collaboration with Scottish Fire and Rescue service and two other VR simulations developed internally at the University of the West of Scotland. However, due to the Covid19 pandemic, difficulties with accessing VR equipment prevented use of the other three VR simulations. As the oil-rig training simulation developed by Swagelok Scotland replicated their classroom-based authentic assessment methodologies for trainees demonstrating their skills and knowledge in a real-world scenario, it provided the perfect scenario to evaluate the framework, allowing the trainees at Swagelok Scotland to reflect on their experiencing using the framework and compare the approach to during classroom-based assessments. Subsequently, the findings from the evaluation would determine the effectiveness, usability and usefulness of the tools, features and components of the framework for performing authentic assessment in VR, directly from a sample of skilled assessors.

5.2.2 Embedding the Framework Into Simulations

As the framework supports any simulation developed using the Unity game engine, it is possible to embed the framework into new and existing simulations. After importing the framework assets, the framework prefabs can be dropped into the Unity game engine scene with all necessary variables assigned. Embedding the framework into a simulation requires a developer to assign variable data for non-deterministic and user input. Although the framework attempts to capture all non-deterministic data by scanning the variables within the Unity scene(s), some non-standard variable information assigned by a developer may not be identified during this process, requiring technical involvement by a developer. After a technically skilled user has assigned all necessary variable data fields, the framework will automatically capture the trainee performance (user input and non-deterministic) data at runtime when the simulation is launched. To capture all remaining data, the framework uses a processed recording layer to simulate the presence of a trainee operating the simulation. This processed recording layer is automatically performed once an end-user assigns the captured data-sets within the framework interface, using the local file directory to indicate the location of folders where the lightweight recording data is stored. This approach is possible due to the use of a novel dual recording methodology to capture the data.

5.3 Dual Recording Methodology Overview

To capture trainee performances in room-scale VR training in sync with the run-time simulation so it can be replayed non-linearly, the data capturing problem was divided into two recording components: lightweight recording, and processed recording (see Figure 5.5), which are designed to overcome the previously discussed concerns and problems with digital game recording, such as Unreal Engines Replay System (see Section 2.7.9) and Unity's Gameobject Recorder (see Section 2.7.6). In VR a minimum frame-rate of 75 frames per second (FPS) is required to prevent motion sickness from affecting the user (Jerald, 2016) i.e. each second, a simulation should be rendered at the display refresh rate of 75Hz without introducing lag in the movements (especially head rotation) and interactions of the trainee with the virtual environment keeping a 'smooth' lag free simulation (Oculus, 2021). Because recording game data during run-time simulation impacts the expected performance of a game or simulation (as discussed in Section 2.7.6), the recording process requires the computer hardware to complete additional recording work while maintaining system responsiveness for the VR trainee. Using current methods for capturing data to fully replay a VR simulation has significant impact on the performance of the simulation, causing 'system-lag' and inducing motion sickness (Jerald, 2016). Because of these reasons current data recording methods do not allow the capture of all simulation data at once, in sync with run-time simulation. The novel solution proposed in the framework allows the simulation to maintain the recommend frame-rate 75Hz at run-time, while capturing a complete data-set of the VR trainee simulation for assessment.

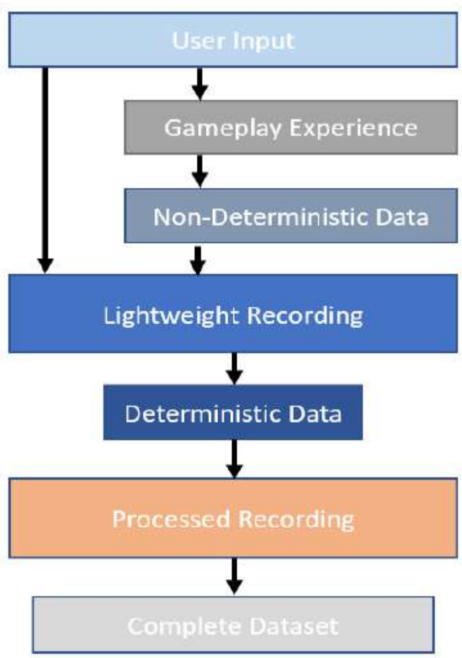


Figure 5.5: Overview of the data accessed by lightweight recording and processed recording. The output generates a complete data-set of all simulation data.

Lightweight recording operates by capturing user input and non-deterministic data, such as AI or physic elements which, due to their design or numerical errors, do not behave in exactly the same way between simulation executions, ignoring all other simulation information, capturing sufficient data that enables a linear reconstruction of the trainee performance, i.e. the captured performance can be consistently reconstructed from the beginning of the trainee performance. Next, processed recording, captures all remaining simulation data that is deterministic, filling the gaps of information not previously recorded by the lightweight recording, allowing the trainee performance be reconstructed non-linearly (see Figure 5.6). The remainder of this section will explain these two components in detail.

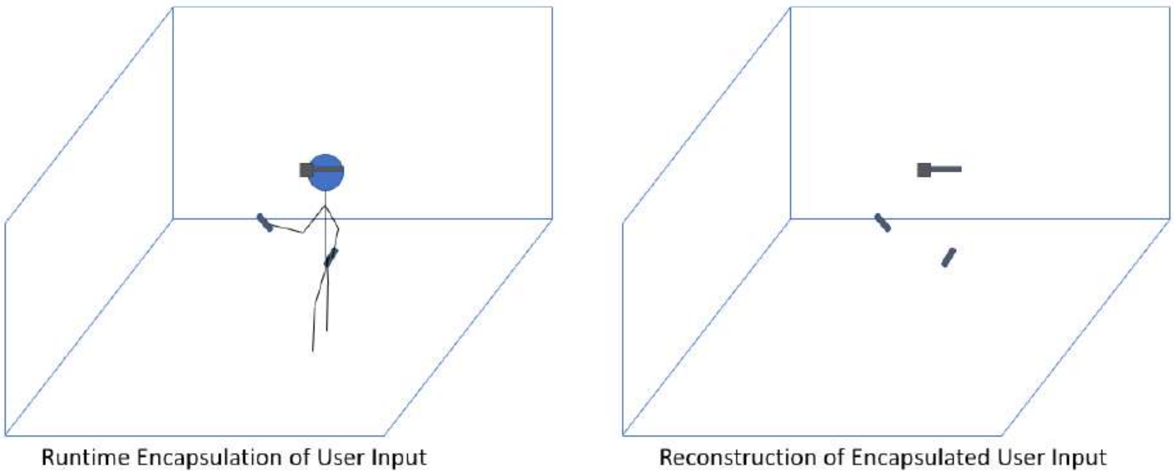


Figure 5.6: (Left) Trainee using VR equipment at run-time, with data capturing the user input and non-deterministic data. (Right) Reconstruction of the virtual reality trainee input to reproduce the input and actions, imitating the experience of a trainee being present and interacting with the simulation.

Lightweight recording operates concurrently with the trainee during the simulation to record user input and non-deterministic data, ignoring all game data that is deterministic, regardless of its importance to simulation or assessment. By capturing sufficient data to satisfy a linear reconstruction process, all deterministic data can be reliably reproduced exactly as experienced by the trainee when reconstructing the simulation. This approach keeps performance impact minimal, as only the necessary minimum amount of data that enables linear reconstruction is captured at run-time, reducing the computing power required to operate the recording process. This reduced requirement for computing power makes the methodology viable for a wide range of virtual reality equipment, including portable and room-scale (see Section 2.9). When a trainee ends their simulation performance, lightweight recording generates and stores a log of all user input and non-deterministic data modifications captured from the simulation in a database (see Section 6.2).

The data captured from lightweight recording alone is insufficient for the data to be useful for observation or assessment by an assessor, as lightweight recording only enables linear reconstructed from the start of the simulation, with no control over the observation time or in-depth assessment of the trainee performance (see Section 4.2). To acquire all information and data from the simulation, and to best present the trainee performance for assessment, a processed recording component is used to complete the recording process.

Processed recording expands upon the data captured from the lightweight recording component, recording all deterministic data modifications from the trainee performance (see Figure 5.7). This enables additional information of the training performance to be added to the data collection and used to provide assisted assessment features required by assessors (see Section 4.3). While lightweight recording, which is embedded into the simulation experience and operates concurrently with the trainee operating the simulation, processed recording operates offline, after the lightweight recording log has been generated. Processed recording operates by reconstructing the linear simulation using the lightweight recording captured data, and recording all modifications of deterministic data, generating a complete history of all data throughout the training simulation experience.

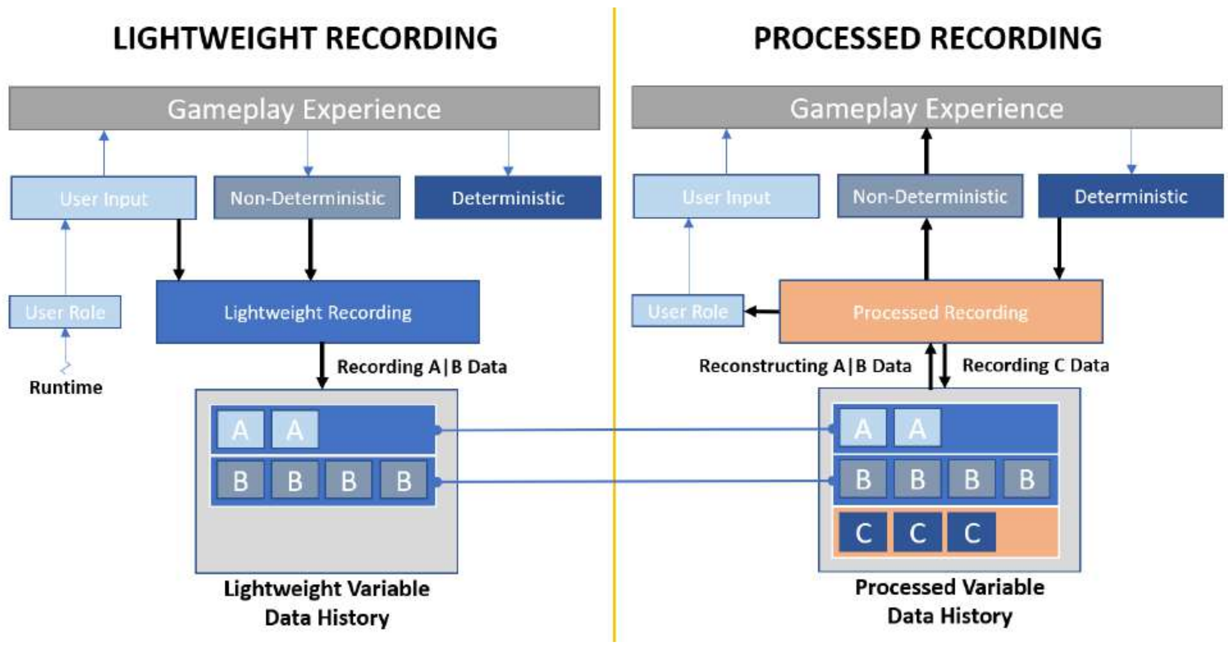


Figure 5.7: The two recording components (lightweight and processed) data saving and reconstruction operations. Black arrows indicate alterations to the structure of the simulation experience (saving or reconstructing data). Blue arrows indicate normal flow of simulation data operations.

With a complete capture of the trainee performance from the dual-recording methodology, all information of the trainee performance can be presented for assessment, providing a complete in-depth overview of the trainee performance, as will be demonstrated with the assisted assessment tools and features presented in Section 6.5. With this information, all simulation events, actions and individual variable modifications can be made available for an assessor to instantly analyse, without having to rely on linearly watching the trainee performance and waiting for actions or events to take place naturally. Furthermore, the using a dual-recording methodology, features such as rewinding, fast-forwarding and time-skipping of the trainee performance are made possible, as well as searching for individual variable modifications and identifying timestamps of interest.

Data is separated into three categories as part of the dual-recording methodology: *user input*, *non-deterministic data* and *deterministic data*, all of which are variables of data within a simulation.

Understanding these three data categories is important to understand the operation of the dual-recording methodology. *User input* refers to as any form of interaction a trainee makes within a simulation, such as a button press or movement of a tracked VR controller. *Non-deterministic* data has an element of unpredictability applied to it. As such non-deterministic data is calculated in a way that the data is not consistent even when provided with the same initial starting conditions. For instance, due to numerical errors, game-engine physic computations may generate similar but non-identical output data when given the same initial conditions. The final form of data is *deterministic data*, which refers to data that will remain constant given the same conditions within a simulation. Unlike non-deterministic, given the same starting conditions, deterministic data will consistently reproduce the same outcomes.

In summary, the dual-recording methodology utilises a two recording components; lightweight recording that operates at run-time with a trainee conducting the virtual training simulation, followed by processed recording which operates offline. Combined, these recording components generate a complete recording of the entire simulation training experience, enabling non-linear observation, with assisted assessment functionality. This recording methodology is only possible due to the architecture for handling data within the framework, which enables all variable data within simulations to be monitored, recorded and reconstructed, without any overhead. To understand how the dual-recording methodology functions, it is important to first have a look at how variable data is handled within C-sharp.

5.3.1 Handling Data in C-sharp

As discussed, the unity game engine is the primary engine used to develop for VR applications (Latham et al., 2019). As such, the design of the components use C-sharp terminology, as C-sharp is the primary programming language for developing with the Unity game engine.

Before discussing the operations for recording and reconstructing data, it is important to clarify how data is managed and accessed by discussing variable handling, casting and reflection. The following explanations are demonstrated within the C-sharp programming language, which is object-oriented (Microsoft, 2021). Within C-sharp, all variable data types derive from the base type of *object* (Clark and Sanders, 2011; Microsoft, 2021), as shown in Figure 5.8. Because C-sharp is a unified type system, all variable types, including predefined and user-defined, reference types and value types, inherit directly or indirectly from type *object* (Microsoft, 2021) (see Figure 5.8). C-sharp includes several predefined variable types, i.e. integer, boolean and string (Microsoft, 2021), along with a variation of custom types (Microsoft, 2021). Values of any type, predefined or custom, can be assigned to type *object* variables (Microsoft, 2021).

To enable dynamic monitoring, handling, recording and reconstructing of data, the framework manages data using type *object* during all operations when managing data internally within the framework. By managing data with type *object*, the framework is able to re-use functions and

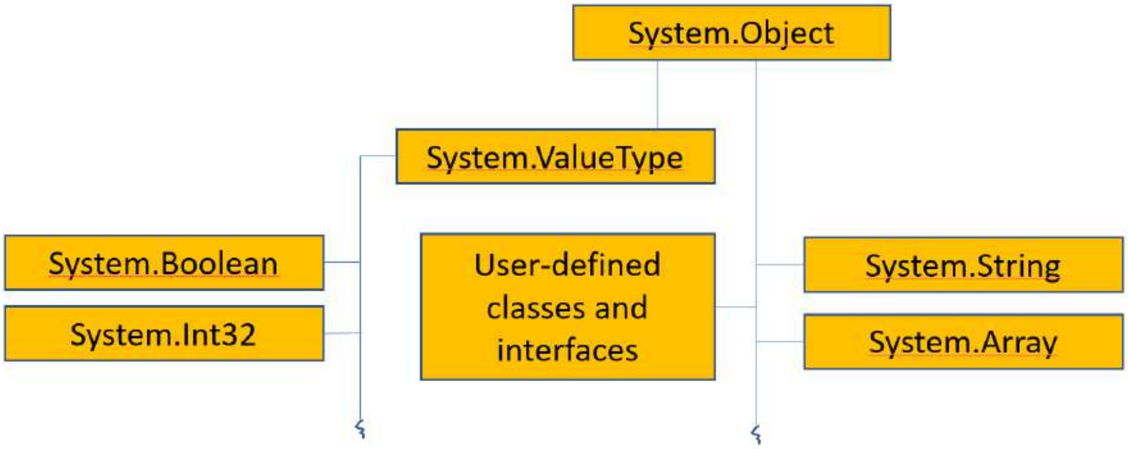


Figure 5.8: In this figure, the hierarchy of data is presented, showing that all other variables inherit from type *object*. Adapted from (Microsoft, 2021).

methods, allowing all variable data to be compatible with predefined *object* input and output parameters within methods and functions. This requires data to be cast from their defined type to type *object*. This is referred to as boxing a variable (Microsoft, 2021). When variables are cast to type *object*, it is boxed (Microsoft, 2021). The variable value is unboxed when an *object* variable is converted to an explicit variable type (Microsoft, 2021).

There are two types of casting, implicit and explicit (Clark and Sanders, 2011). Implicit conversion enables the compiler to automatically cast data to predefined compatible types (Clark and Sanders, 2011). Explicit conversion is referred to casting, as it requires a specified type supplied within parenthesis to explicitly cast data from one type to another. As such, the framework boxes variables by using implicit casting to cast variables to type *object* for all functions and operations throughout the internal logic of framework. *Object* variables are unboxed for any use purpose, such as saving or reconstructing the data. When saving, the variable data is implicitly cast to type string for universal compatibility with Json and XML file extensions. During reconstruction, the variable type is required to be unboxed to its specified variable type used for the game engine. Outside of these two instances, variable data is handled and manged as a boxed *object* type.

In summary, the framework incorporates these technical implementations to manage variables as type *object*, which requires implicit casting to cast variables to type *object* during recording (Microsoft, 2021). During reconstruction explicit casting is required to casts object variables back to their original variable type (Microsoft, 2021), this is achieved using a library system that configures the data (see Section 5.3.3). However, first it is important to clarify how data properties and fields are accessed and modified.

5.3.2 Accessing Variable Values

Reflection provides the ability to dynamically access properties and fields within an existing *object*, or invoke methods (Microsoft, 2021). Reflection enables properties and fields to be read or overwritten, through get and set commands (Microsoft, 2021). This functionality provides the ability for the framework to record and reconstruct data dynamically. During recording, the framework monitors and saves variable data by monitoring the current value accessed through system reflection, and comparing it against the previous value. During reconstruction, variable information is overwritten using system reflection set command, replacing the current field or property of the variable with the previously recorded value. As the framework relies on system reflection to access, read and write variable values using variable information (name), the data-set generated by the framework explicitly matches the game engine hierarchy, as discussed in Section 5.3.4.

Each individual variable being recorded is assigned an individual-recording instance to monitor the variable value for changes. Every frame the framework monitors the previous value of the variable with the current value, by accessing the value field of the variable being monitored using reflection (Microsoft, 2021). If the previous cached value and current value of the variable are different, the framework registers that variable data has been modified and begins recording the modification (see Figure 5.9). This operation is identical for all variables being monitored. As discussed in Section 5.3.1, the variable value is saved to a database using type string. As such, the framework performs a implicit casting to unbox and cast the object to variable of type string (Microsoft, 2021) (see Figure 5.9).

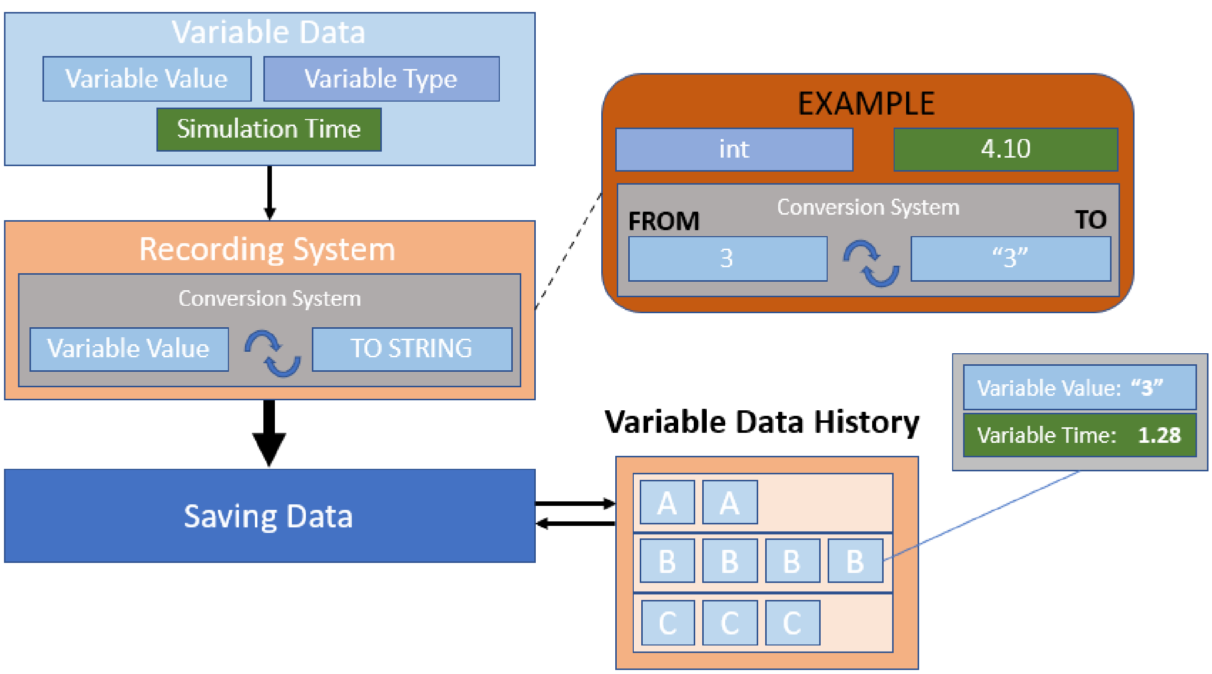


Figure 5.9: The saving process including the conversion of variable data to type string (text), with the modified value saved and included in the history of modifications of that variable.

However, to reconstruct variable data stored in the database, explicit casting is required to cast variables back to their original type required by the game engine (Microsoft, 2021). To address this issue, a library system was created to configure variable types to enable dynamic casting of data during reconstruction operations.

5.3.3 Library System

As discussed, data is monitored and handled using the type *object* identifier, as it is universally compatible for all functions (Microsoft, 2021). When saving, the data is implicitly cast to type string, converting the variable data from its original type to string so it can be saved to a database for storage during recording. This is possible as string is a predefined type and does not require special syntax (Microsoft, 2021). However, to cast saved data from the database back to the original variable type, explicit casting is required, as the reconstruction of the data requires the types to be explicitly defined for the compiler (Microsoft, 2021). As such, casting data from object to a explicit type requires the variable type to be defined within parenthesis (Microsoft, 2021).

To streamline this process, a library system was created for the framework that supports the majority of predefined variable types and allows easy extension to other custom types. This library system was developed to provide the ability to dynamical cast data by positioning the casting process into a single section of the framework (see Figure 5.10), which meant that explicit mentioning of variable types requires only manual input by the developer once. As presented in Figure 5.11, the library system allows explicit casting of variable types that are known within the library system.

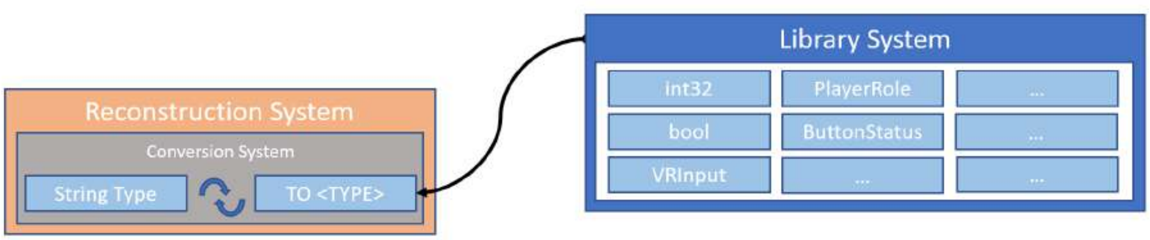


Figure 5.10: A demonstration of the library system defining explicit casting (Microsoft, 2021) information for the framework to dynamically cast data during reconstruction.

Listing 5.1 shows an extract from the library system in pseudocode, demonstrating how the library system casts variable data when reconstructing simulation data using the variableType identifier previously presented in Figures 5.10 and 5.11.

```

If variableType equals Bool
    return variabledata as type Bool
If variableType equals String
    return variabledata as type String
If variableType equals Float
```

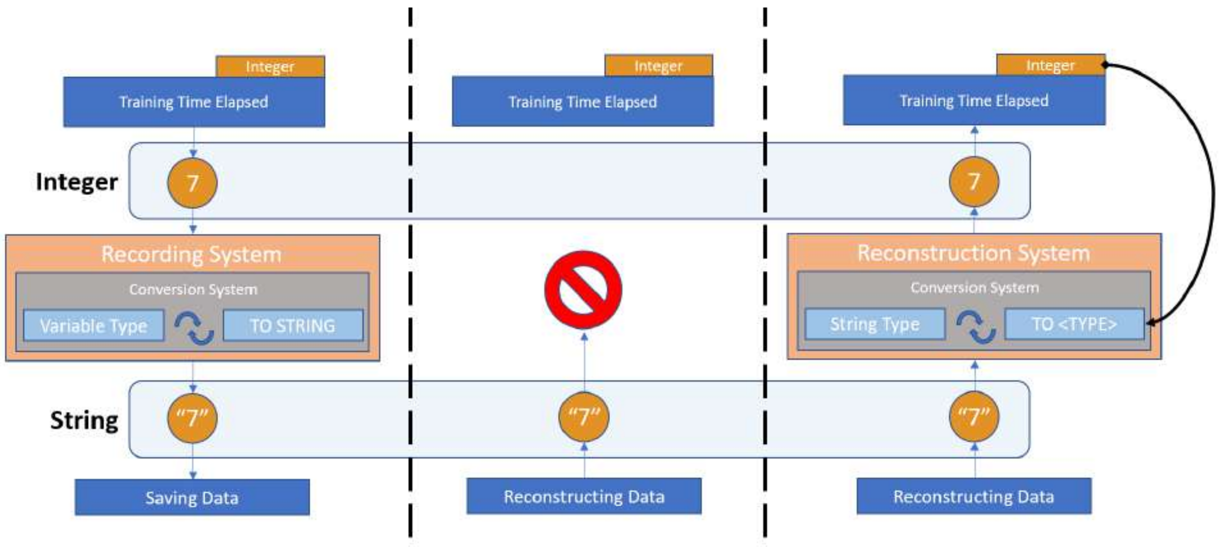


Figure 5.11: Demonstration of the process a variable goes through being recorded and reconstructed. Without a casting (conversion), the data is incompatible for saving or being reconstructed. (Left) the integer time value is converted to string and saved. (Middle) the saved data attempts to reconstruct the data without casting the variable from string back to integer, causing the simulation to crash. (Right) shows the same reconstruction example, but this time with a casting system, enabling successful reconstruction of the variable data.

```
    return variabledata as type Float

return null and alert developer if no suitable variable type is found
```

Listing 5.1: Shortened pseudocode of the casting system found within the library system.

5.3.4 Recording Hierarchy

In order to record data using the dual-recording methodology, a database structure was created to keep the data in a hierarchical format that enabled values be added and removed as necessary. This allows lightweight recording to create the initial database to capture all user input and non-deterministic data. Processed recording then uses this database to reconstruct the simulation, and populate it with the remaining data (deterministic data).

The recording of data operates by creating a hierarchy of object information for each individual object, this is split up into specific elements attached to each unique object within a simulation. Each object is assigned a unique identifier-key to identify it during recording and reconstruction operations. For instance, 'GameObject A' includes an assigned name and unique identifier-key. Each object within a simulation will contain default properties that are always present, including activation state, position, rotation, and scale. Beyond these default properties, each object can contain additional data elements assigned as components, being attached to the individual object. Embedded in these components are additional variable data which are used to drive the properties of the customised object, for example the movement speed of a car within a driving simulation. A recording hierarchy is created that stores data in a hierarchy, capturing a history of each individual variable attached to a single unique object (see Figure 5.12).

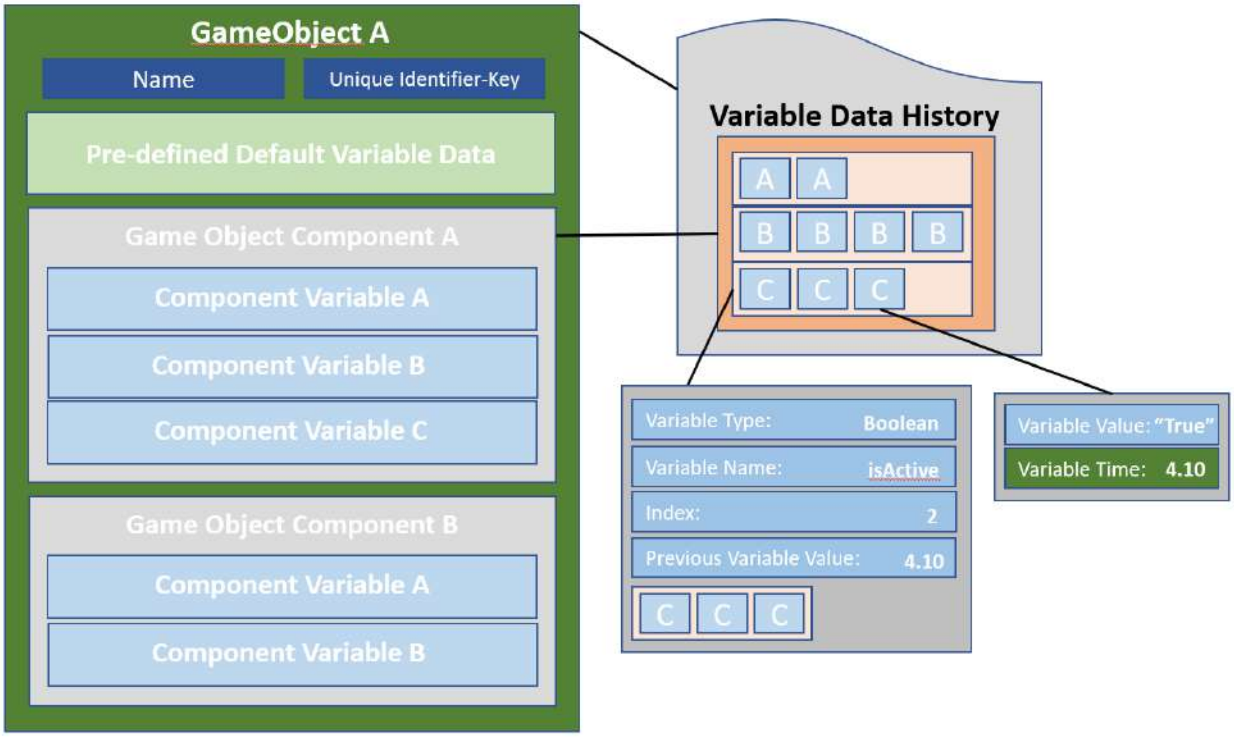


Figure 5.12: Data recording structure of object information.

Default data: Activation State (boolean), Position (Vector3), Rotation (Quaternion) and Scale (Vector3) are the only variables that are stored using their real-variable types for the sake of convenience when debugging and testing, bypassing the casting or converting systems of all other data. Since default data is predefined data and exists for all objects in simulation, saving this data as the variable's real type, reduces the processing on the system during recording and reconstruction. Because default variables are considered the most likely variables to be modified, handling and storing this data using the default variable type can improve performance of the processed recording component. While default data is saved using its own variable types, all other variable data is handled using string variable type. Because all other variable data that exists during simulation is unknown by the system, a universal approach to handling and saving the data is necessary. Consequently, careful consideration for handling and accessing variable data is required to enable dynamical capture and reconstruction variable data.

With this approach to recording data, lightweight recording creates the hierarchy of the recording structure. Data captured by lightweight recording at run-time is marked as red in Figure 5.13, highlighting the data has non-deterministic elements or characteristics. Alternatively, data captured during processed recording is designed to capture all remaining deterministic data, marked as blue in Figure 5.13. This hierarchy produces a complete data-set of the simulation performance, including the modifications for of each individual object within a simulation, and furthermore, the modification of each individual data element embedded on the single object, creating a full history of all variable data modifications. The captured information combines to provide a rich detailed insight into the trainee performance that is unmatched by any single recording approach.

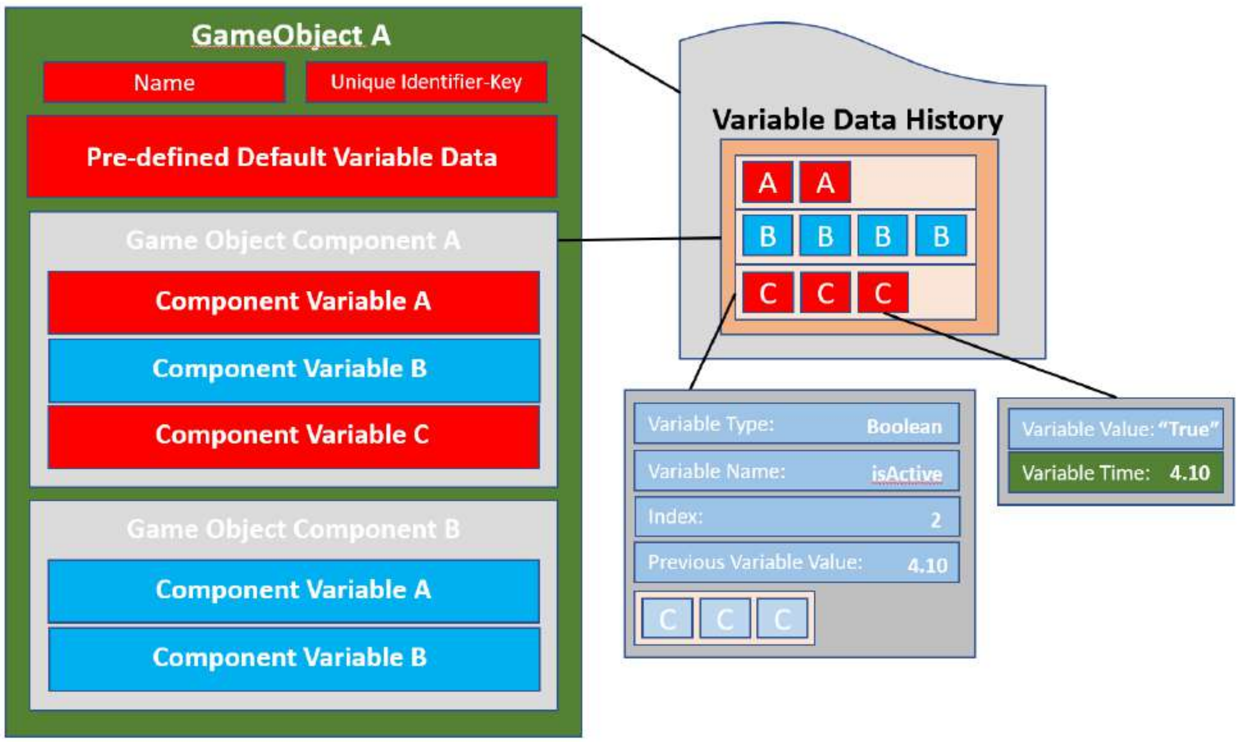


Figure 5.13: An example of the data recording structure showing the classification of *non-deterministic* (RED) and *deterministic* (BLUE) data captured for an object within the simulation environment.

5.3.5 Data Generation Output

With the dual-recording methodology, there is no limit to how much data can be captured during the processed recording component. By using a dual-recording methodology, all variable data from a simulation is recorded and available for in-depth analysis during assessment (see Figure 5.14).

All variables from the trainee performance are timestamped against their recording instance, this makes reconstructions of the experience more usable and useful for observations or assessment, as will be discussed in sections 6.3 and 6.5. Because the data is stored in a database, the history of data modification is portable and can be easily transferred from one device to another, or to a server (Howie and Gilardi, 2020).

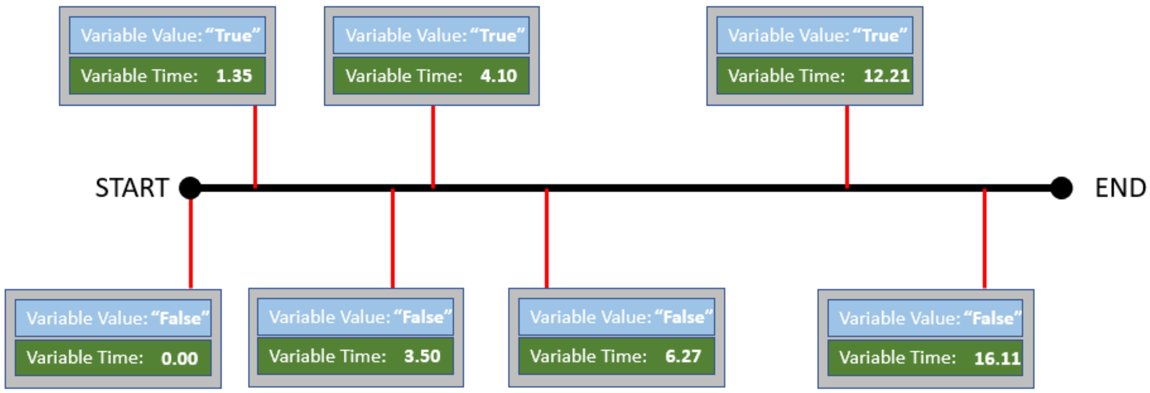


Figure 5.14: An example of data from a variable captured by the dual-recording methodology. The modifications with time-stamps are used to provide assessors with a complete history of each modification of a variable throughout a trainee performance.

5.4 Lightweight Recording

As stated in Section 5.3 the recording operations for capturing data is split into two recording components, this section focuses on the initial component, lightweight recording.

The lightweight recording component is designed to capture all trainee interaction by monitoring and recording all modifications of input controls, such as movement, rotation and controllers inputs, along with other tracking devices, such as eye-tracking and audio input (see Figure 5.6).

Beyond user input, lightweight recording captures all non-deterministic data from the simulation, recording data when any modifications of the monitored values are detected. These recordings are conducted in sync with the user input, logging the data with timestamps to create a timeline for when each variable was modified, generating a history of each variables value modification.

Because consistent reconstruction of data is required to reproduce the simulation for a consistent observation and assessment, all non-deterministic data has to be recorded and reconstructed by the framework. For example, if a simulation relies on a random fault from a database at the start of a training simulation, lightweight recording must capture the configuration of fault at run-time to keep a record of it. This means the reconstruction of data can reproduce the same simulation fault and outcomes, ensuring the simulation data that is reconstructed is consistent and in sync with the actual performance of a trainee.

Because lightweight recording operates at run-time, it is essential that its impact on the simulation is minimal, as recording too much data at once will severely impact the run-time performance of the simulation. By capturing only the necessary minimum amount of data (user input and non-deterministic data) that enables a consistent linear reconstruction, the impact of recording at run-time negligible as only a small amount of data is captured. This satisfies one of the design challenges for capturing training simulation performances, as the framework is viable for use in low-end hardware, such as portable VR equipment, along with low performance computers without having any significant impact on the responsiveness and interactivity of the hardware. This minimal impact on run-time performance is possible by the use of recording instances to captures the modifications of data.

5.4.1 Recording Instance

A recording instance is the term used for indicating a variable being monitored has been recorded. As discussed in Section 5.3.2, when a variable being monitored value changes, the modification is saved by recording an instance of the value changing.

As the framework monitors each variable individually, the framework assigns each variable within the database two separate lists for time and value modifications (see Section 5.3.4). When a recording instance is triggered, the variable value and time of modification (which is managed internally within the framework) is saved to the database. By recording modifications as individual variable instances, it is easier to record and reconstruct individual data as necessary, rather than all data being saved or reconstructed at once (Howie and Gilardi, 2019; Howie and Gilardi, 2020).

Throughout this framework, the term *recording instance* will be used when referring to an instance where a variable value has been saved by either recording component (lightweight or processed recording).

5.4.2 Lightweight Recording Functional Operation

The lightweight recording component is designed to allow for unconstrained and customised recording of user input that can be modified to suit the needs of the hardware configuration and simulation requirements. Because each input device has different forms of functionality and usability, the component can dynamically be extended to suit the specific needs of the simulation experience being recorded. For example, a computer keyboard, VR controller, or both can be recorded simultaneously. The component monitors all binded input from these devices and keeps track of any modification, and triggering a recording instance of the new data when it is detected. Binded input can be added or removed as necessary for the expected user input.

Input and non-deterministic data captured by lightweight recording is assigned at the start of the recording process. It is partly automatic, with the system automatically detecting physics and other non-deterministic data for capture using embedded algorithms that detected specific components for randomness elements. User input is configured to capture all current VR hardware, including button presses and motions that can be accessed by the development kit. The input and custom data, can be modified or added as necessary to suit the needs of the simulation. For example, a new VR controller can be mapped to input data which will be captured by the recording system. Prior to a simulation starting, all user input, non-deterministic and custom data is assigned for monitoring by the system. Additional input hardware, such as eye-tracking, is easily embedded into the system by assigning the variable input as user input. This structure enables lightweight recording to be robust but lightweight, offering a targeted approach to capturing data that can be modified as necessary.

The recording system relies on the architecture of the framework for handling variables discussed in Section 5.3. During recording, the recording components keeps track of all monitored data. At the end of each frame, the simulation monitors the current value of data and compares it to the previously recorded value storing them in an active database. If no previous value exists, the current value is assigned as the previous value. When the current and previous values are different, the recording component initiates a recording instance to record the modification. This recording operation updates the previous value to match the new value of the data, logging the modification in the database along with a timestamp of current simulation time (see Section 5.4.1). Each variable modification recorded creates a history of the variable values throughout simulation starting from the initial value when simulation begins until the recording terminates when a simulation is stopped. To maintain frame-rate and have negligible impact on the simulation performance, the recording operation only activates when current value and previous value of monitored data are different. Figure 5.15 shows how variable data is recorded, showing the process each individual variable being monitored goes through each frame.

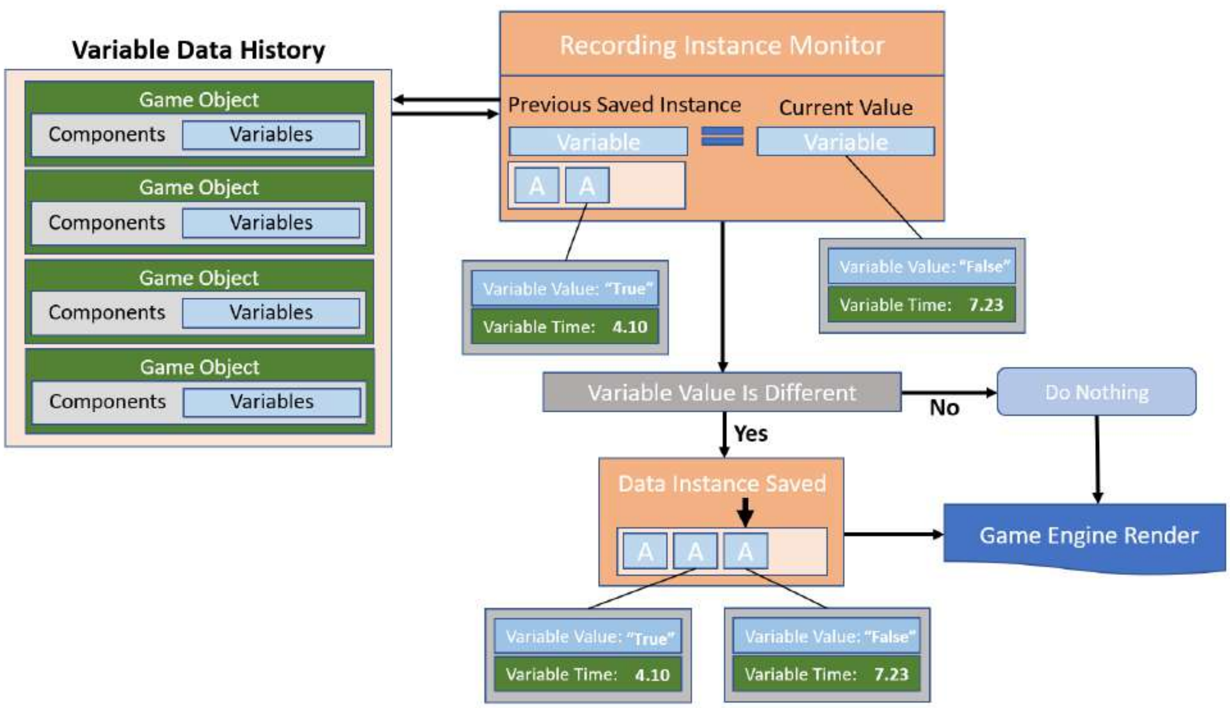


Figure 5.15: In this figure, the system monitors the current value of data against the previous value. As it is different (current value: true, previous value: false), the previous value is updated to equal the current value. The data is then saved to the database along with the time instance of the modification.

With the data captured from lightweight recording, the trainee performance can be reproduced linearly, from the start of the simulation until the end of the recorded data. Data is reconstructed when the timestamp of the recording instance matches the current reconstruction time. The resulting linear reconstruction reproduces the simulation exactly as experienced by the trainee.

However, while the data captured from lightweight recording can enable direct observation of trainee performance, it does not satisfy the requirements identified in Section 4.2 for conducting authentic assessment and contextual feedback, as assessors require the ability to rewind or skip sections of the trainee performance. Linear reconstruction can only provide a replication of recorded data from the beginning of the simulation and cannot be skipped ahead or reversed, as each recorded data needs to be recorded in sequence to consistently reconstruct the deterministic data of the simulation. As such, assessors are unable to identify key-points of trainee performance data, as only a limited amounted of data has been captured by the lightweight recording. This forces assessors to linearly watch through a simulation until key-points of trainee performances are reproduced. To solve this issue, the framework uses a second recording component, processed recording, to linearly reconstruct the trainee performance and capture all remaining deterministic data offline, providing a complete capture of all trainee performance data.

This approach allows the observation, assessment and feedback components, tools, and features to meet the requirements identified from the functional requirement collection study (see Section 4.2) with assessors presented in Section 4.1.

5.5 Processed Recording

Processed recording uses the data captured by the lightweight recording component as the basis for reconstruction. Data captured by lightweight recording reproduces the user input and non-deterministic data linearly. This linear reconstruction reproduces the outcomes and modifications for deterministic data. As such, processed recording captures all deterministic data, filling in the gaps left by lightweight recording in the recording hierarchy (see Section 5.3.4).

By capturing all outstanding simulation data, subsequent reconstructions of the trainee performance, such as for observation (see Section 6.3) or assessment (see Section 6.5), can use a non-linear reconstruction methodology. Non-linear reconstruction keeps user input, non-deterministic and deterministic data in sync, allowing additional functionality, such as fast forwarding, rewinding data, and time skipping, which is necessary for the requirements identified in Section 4.2.

Unlike lightweight recording which operated at run-time, processed recording operates offline, after the simulation has concluded. Due to the portability of the data generated by the lightweight recording, processed recording can be conducted on alternative hardware or devices. For instance, lightweight recording data captured on a portable standalone VR headset can be transmitted to a server for processed recording. Once processed recording has been conducted, the data can be returned back to the standalone headset for observation or assessment.

Although processed recordings role is to generate a complete data capture of all variable data throughout a simulation experience, it is not always possible to capture every element of data. In some instances, data based on internal game engine logic that is write and read protected cannot be captured by processed recording. Security protected variable data is also ignored as they are likely to be irrelevant for the reconstruction of a trainee performance. Both the lightweight and processed recording components use same recording operations for monitoring, accessing and saving data.

5.5.1 Processed Recording Operations

Processed recording operates by configuring the simulation session to operate with the linear reconstruction methodology. This linear reconstruction reproduces the trainee performance using the data captured by the lightweight recording component. Processed recording scans all variable information throughout the game experience and checks if the data exists in the lightweight recording data hierarchy (see Section 5.3.4). If the data does not exist in the database hierarchy, the data is marked for monitoring. Figure 5.16 presents the processed recording operation, showing the reconstruction of simulation to capture non-deterministic data.

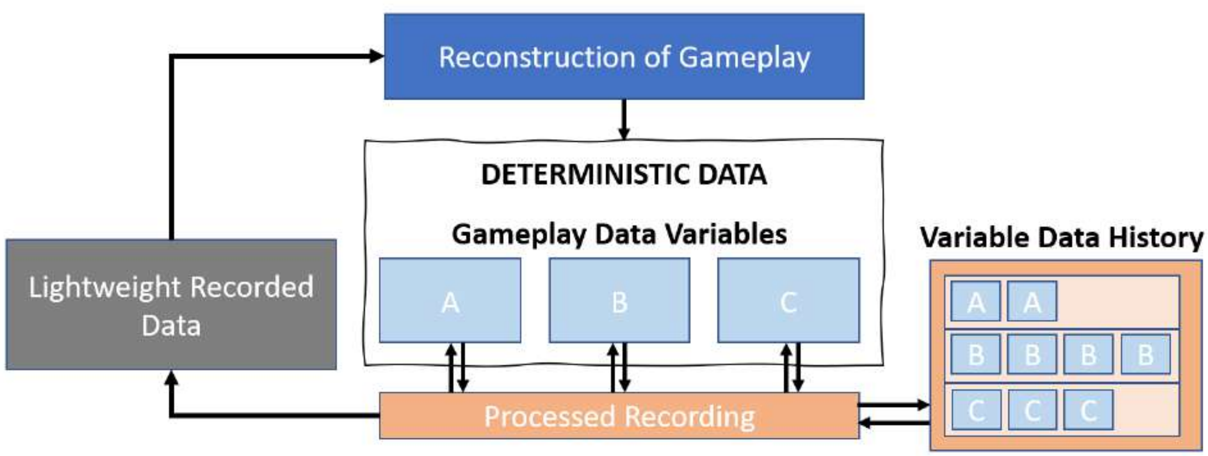


Figure 5.16: Processed recording reconstructing the linear simulation experience using data captured form lightweight recording. During reconstruction, all outstanding variable data is monitored for changes to capture all deterministic data modifications and store them within the database.

Because the processed recording operation is required for each individual trainee data-set, it is designed to operate independent and automatically, allowing one or multiple trainee performances captured from lightweight recording component to be queued and processed with no additional work or input required by a developer, trainee or assessor.

The processed recording operation automatically outputs and prepares a complete data-file of the trainee performance which includes the modification of all data within a simulation for a trainee performance, including user input, non-deterministic and deterministic data. This is stored within the database. These design choices streamline the process of capturing data as part of the framework, minimising the workload required for assessors. In the case of remote training, it would be envisioned that trainee simulation data would be uploaded to a server for storage, with the server automatically processing the data, preparing it for assessment by an assessor as demonstrated in Figure 5.17.

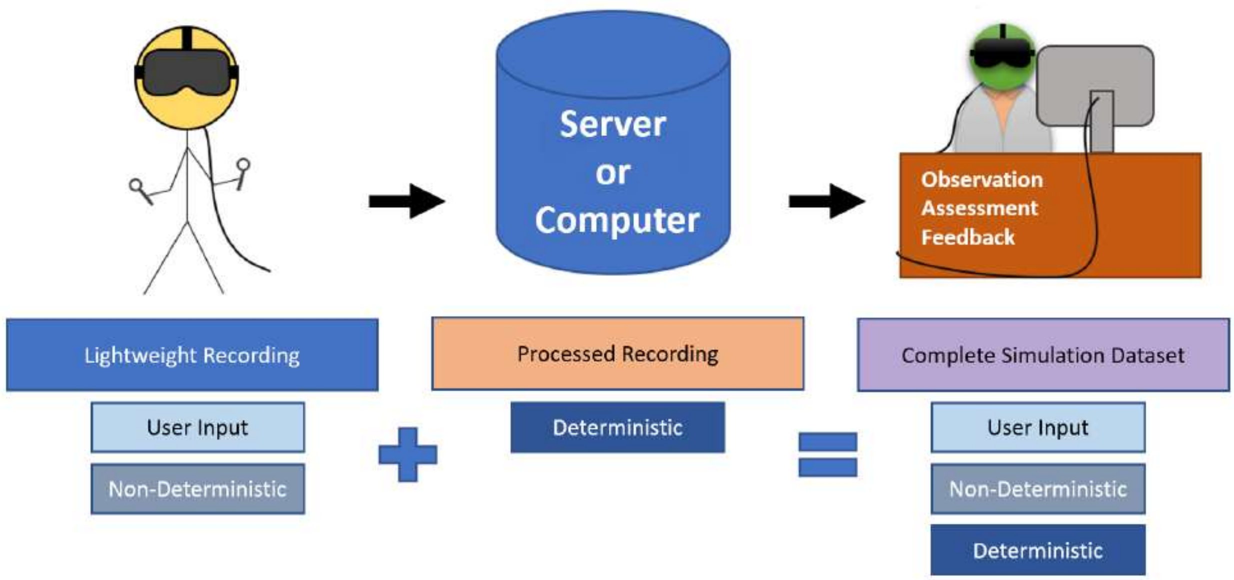


Figure 5.17: A demonstration of Lightweight Recording capturing User Input and Non-deterministic data, followed by Processed Recording conducted on a server or a computer, producing a complete data-set of the trainee simulation.

Listing 5.2 shows the pseudocode for the processed recording operation, detailing the steps to acquire all information, identify duplicated names, and demonstrating the characteristic requirements used to determine if an object is suitable for recording.

```
if ProcessedRecording equals true AND ProcessingMethod equals all AND current status of recording is not equal recording
    Set Quality Low
    Find all gameobjects previously saved and iterate through them, informing them to ignore any Processed recording calls

    Find all gameobjects within the scene using _transform
    iterate through all found _transfom gameobjects
        set found to false
        iterate through all previous saved gameobjects
            if _transform name equals any previously saved gameobject name
            set found to true

        initialise repeated object count as zero
            iterate through all objects in _transform
                if _transform name is repeated
                    increase repeated object count

            if repeated object count equal to or more than 2
                alert developer
                break iteration

        if gameobject is eligible to save AND repeated object count is equal to one AND found is false AND _transform has not already been configured for Lightweight Recording or Processed Recording
            add recording components to _transform
            configure _transform for recording
```

Listing 5.2: Pseudocode of the data being configured for Processed Recording.

In summary, processed recording expands upon the lightweight recording simulation data, by capturing all outstanding data which is deterministic. This operation generates a complete data-set of a trainee performance, which is stored in the database. As such, the database has access to all modifications of simulation data during a trainee performance, allowing a timeline history of modifications to be provided for assisted assessment features, such as the summative performance indicators (see Section 6.5.1) and an augmented timeline (see Section 6.5.2). Furthermore, with all data captured from a trainee performance, the framework can use a non-linear reconstruction methodology, allowing assessors to rewind, fast-forward and time-skip their observation or assessment of trainee performances.

5.6 Chapter Summary

This chapter has presented the patent-pending dual-recording methodology that enables all data within simulations to be captured without impacting the run-time performance of the trainee. This is achieved by categorising variables into *user input*, *non-deterministic* and *deterministic*, to indicate their order of importance for recording and reconstruction. User input and non-deterministic data is captured at run-time using the lightweight recording component, with deterministic data captured within the processed recording component. This combined recording operation allows subsequent reconstructions to use non-linear reconstruction methodology that allows assessors to rewind, fast-forward and time-skip their observation or assessment of trainee performances.

Handling data within C-sharp, the ability to box and unbox variable values using type *object*, the use of reflection to access, read and write variable values and the library system to dynamically cast variable data during reconstruction were discussed, showing their importance for monitoring, saving and reconstructing data. The combination of these designs and features enable all data from trainee performances to be captured without impacting the run-time performance of the simulation. This functionality was presented for the Unity game engine and C-sharp.

With dual-recording capturing all data from a simulation, it is possible to design the tool and features for the observation, assessment and feedback components that meet the functional requirements identifed by the requirement collection study with assessors (see Section 4.2 and 4.3). Chapter 6 presents the design of these tools and features that enable assessors to conduct authentic assessment, and generate contextual feedback of room-scale VR training simulations.

Chapter 6

Framework Components Design

6.1 Introduction

With the dual-recording methodology presented in Chapter 5 capturing all the data in a simulation, it is possible to design the observation, assessment and feedback tools and features that meet the requirements from the study conducted with of assessors from RQ3 (see Section 4.2) and RQ4 (see Section 4.3). These tools and features allow assessors to conduct authentic assessment and generate contextual feedback of room-scale VR training simulations, using the same principles of classroom-based assessment.

This chapter presents the design of the remaining three components of the framework proposed in this work; Observation, Assessment and Feedback. These components combine to create the framework, with each component including the necessary tools and functionality to meet the requirements of assessors (see Section 4.3).

As shown in Figure 6.1, this chapter presents the observation, assessment and feedback components of the novel framework. Following the dual-recording methodology presented in Chapter 5, which presented the lightweight component and processed recording component which capture a complete data-set of the trainee simulation. This chapter concludes the outstanding work necessary to answer RQ5, which asks *"How can findings be compiled to create a framework for a generalised system for observation, assessment and feedback of room-scale virtual reality training simulations?"* (RQ5). This research question is answered by designing the remaining framework components and their accompanying tools and features.

First, it is important to present the database (see Section 6.2), which acts behind-the-scenes, allowing all other components to seamlessly work together. The database stores all data captured from the training simulation by the dual-recording methodology (see Chapter 5) and provides the technical capabilities that enable the observation (see Section 6.3), assessment (see Section 6.5) and feedback (see Section 6.6) components to achieve their role within the framework, providing assessors with the capability of conducting authentic assessment and generating contextual feedback of room-scale VR training simulations.

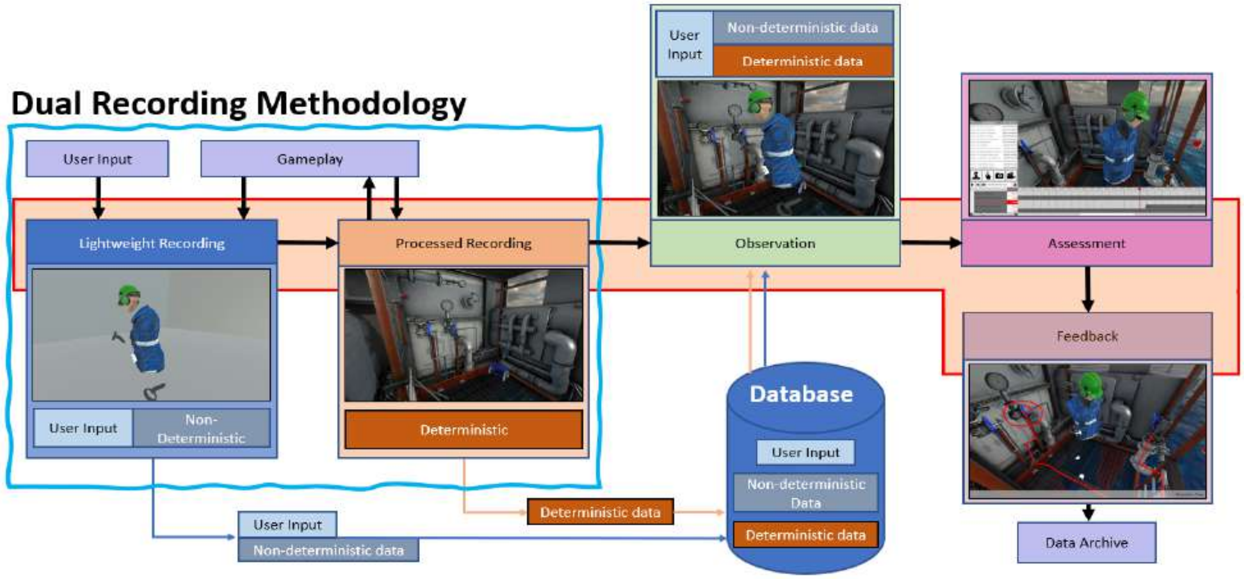


Figure 6.1: As previously presented in Chapter 5, the framework incorporates additional components, tools, and features and to provide the observation, assessment and feedback operations.

6.2 Database

Following the dual-recording methodology discussed in Chapter 5, the data collected from the lightweight and processed recording components is stored in a database, using either JSON or XML formatting. The recording hierarchy for the database was discussed in Section 5.3.4, however it is important to emphasise its role within the framework components for providing the assisted assessment features. As such, the database is designed to enable the observation and assessment components, tools, and features to meet the requirements identified in Section 4.2 from the study with assessors (see Section 4.1).

By formatting the database storage hierarchy to mimic the structure of the Unity game engine, information of the trainee performance is easy to acquire through queries within the framework. By referencing the unique object identifier (or object name), component type (or name), and variable name, any variable captured during recording and stored in the database can be quickly retrieved. This enables the framework to display variable data history and all modifications expected during reconstruction within the augmented timeline (Section 6.5.2), present pass or fail information for summative performance indicators (Section 6.5.1) and keep the simulation in sync during non-linear reconstruction events (Section 6.3).

By replicating the game engine hierarchy within the database, the framework is able to scale and adapt to any simulation. Rather than being restricted to predefined variables (Raymond and Usherwood, 2013), the database structure provides assessors with the opportunity to analyse data that initially may not have been considered important during development. Because the dual-recording methodology captures all data within training simulations, developers or assessors are not required to explicitly define observation or assessment variables that are external to the simulation logic. As such, common assessment and performance values including time completion, number of errors, motion data, objectives completed (Lahanas, Georgiou, and Loukas, 2016; Konakondla, Fong, and Schirmer, 2017; Moore et al., 2020) are all easily accessible by querying the database (Figure 6.2), and require no prior configuration by the developer, as is demonstrated by the summative performance indicators (Section 6.5.1). This means that the database can be used to find and sort data to identify cause and effect relationship values from trainee performances, and eventually moving in the direction of AI and deep learning (as future work), given large amounts of data is made available for in-depth analysis.

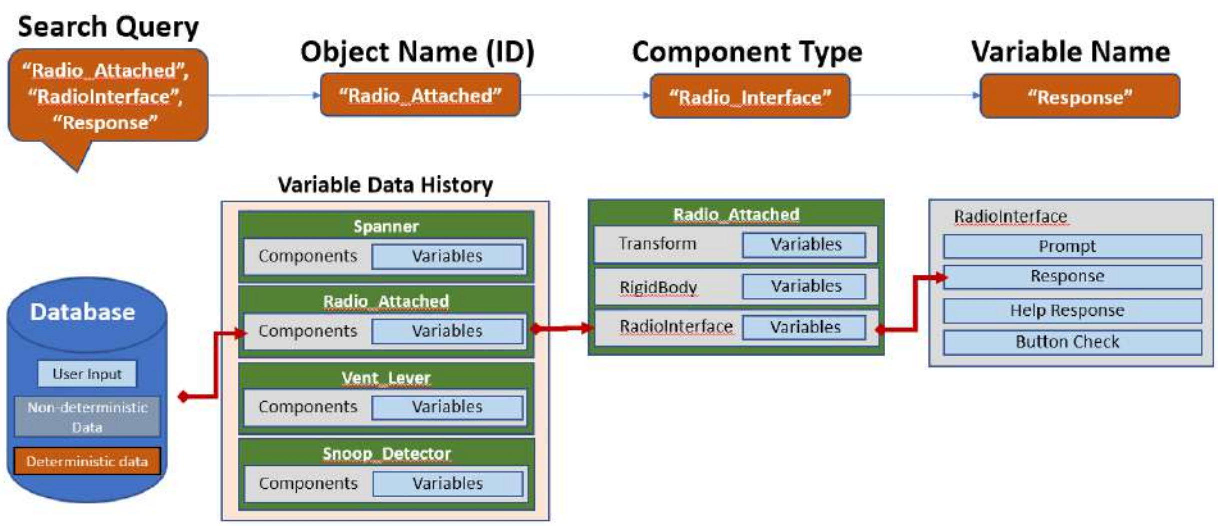


Figure 6.2: A search query for the database which looks for the object *RadioAttached*, in component *RadioInterface* and variable name *Response*.

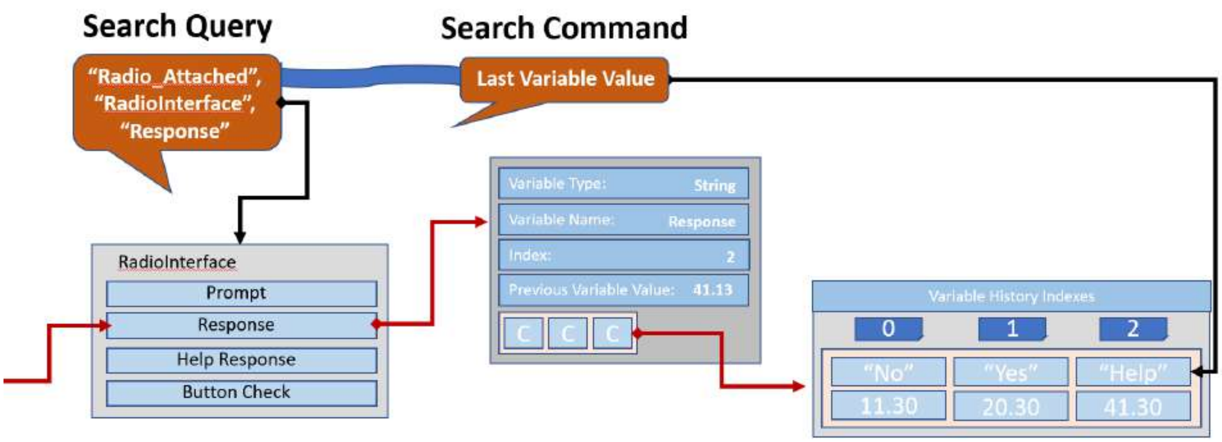


Figure 6.3: Following on from the search query in Figure 6.2, the framework can perform common commands to access elements from the saved variable data history. In this example, the last saved value of the variable is being requested. The last value stored in the variables history is then returned and can be used to identify summative performance indicators (6.5.1) or for presentation on the augmented timeline (see Section 6.5.2).

6.3 Observation

For assessors to make use of the data captured from the simulation, an observation component of the framework is required. This observation component transforms the saved data from a trainee performance into format that enables direct observation of the trainee performance.

The observation component uses the data stored in the database to reconstruct the simulation. Since the database includes all data of a trainee performance in the training simulation, classified as: user input, non-deterministic and deterministic data, a non-linear reconstruction methodology can be used to reproduce the trainee performance. This means, the observation component enables features like rewinding and time-skipping to be used, allowing the observation components tools and features to meet the requirements identified in Section 4.2. This will be discussed in Section 6.3.3.

However, to enable direct observation of the trainee performance, the observation component requires the ability for the assessor to view the recorded data into a format that provides visual representation and reconstruction of the trainee performance, allowing assessors to view trainees applying and demonstrating their knowledge to overcoming objectives and tasks within the simulation. As such, the data stored within the database must be reconstructed into a format that allows direct observation of the trainee performance.

6.3.1 Reconstructing Data

The observation component expands upon the reconstruction architecture used by the processed recording component to reproduce all recorded data. User input is reconstructed to reproduce the trainee movements and actions during their training performance, with deterministic and non-deterministic data reconstructed in sync to create a consistent reconstruction of the trainee performance. Because the data from the recordings replicate the actions and modifications of a trainee, the simulation will react as if a trainee is currently operating the simulation, since the framework mimics the presence of a trainee run-time performance within the simulation. As such, the reconstructed performance will reproduce a consistent reproduction of the trainee performance, allowing after action authentic assessment to be conducted (Hanoun and Nahavandi, 2018; Wiggins, 1990).

Listing 6.1 presents the pseduocode of the implementation for reconstructing data. First the reconstruction determines if the data being reconstructed has reached the end of its stored history index count, followed by a further check to determine if the data has been marked as complete. If so, the function returns and nothing happens for that individual element of data. Next, the current simulation time is measured against the value stored as the next recording instance time. If the recording instance is equal to or lower than the current reconstruction time of the simulation, the data is reconstructed using the *ChangedDataInstance* function. When reconstructing data using the *ChangedDataInstance* function, the framework determines what type the variable is, and then converts the saved string data using the library system back to its original type, and then replaces the current data with the saved data. After the value has been reconstructed, the ForwardRebuild function updates the data to determine if all data for that individual element has been reconstructed. This processes is repeated for every individual element of data that is being reconstructed.

```
---------- ForwardRebuild Function----------
if index is more than the total count of variable data history stored for the variable in the database
    return (since the index is more than the count for saved variable data history)

if isComplete is true the replay mode is constant
    return

if the current simulation time is equal to or more than the next reconstruction point of variable data
    change the data value to reconstruct the next value of data ---> ChangedDataInstance();
    increase the index value
    if index is more than the amount of variable data history
        isComplete is set to true (all data has been reconstructed)

----------  ChangedDataInstance Function ----------
if variable type is of type field
    if data is valid to reconstruct determined by the field type
        vardata equals the data converted from string to the objects variable type
        try
            set vardata to object field value
        catch
            do nothing if setting value fails (alert developer)
else if variable type is type property
    if data is valid to reconstruct determined by the property type
        vardata equals the data converted from string to the objects variable type
        try
            set vardata to object property value
        catch
            do nothing if setting value fails (alert developer)
```

Listing 6.1: Pseudocode of data being reconstructed in the forward reconstruction mode to reproduce the modifications of the simulation in visual format.

The pseudocode is shown in Listing 6.2, shows how the interface locks all interaction with the framework then iterates through the recorded objects to keep all information in sync when skipping areas of the trainee performance. Finally, the framework pauses the simulation for one second to give the game engine time to adapt to the rapid change of variable information before allowing reconstruction to continue for direct observation by assessors.

```

------FastProcessReconstruction------
    lock controls
    set reconstruction direction mode
    set current time as simulation time
    get requested simulation time

    for all saved objects
        for all object components
            for all component variables
                set status of element value at requested simulation time

    set reconstruction time for User Input
 set reconstruction time for Deterministic Data
    set reconstruction time for Non Deterministic Data

    wait one second

    unlock controls
```

Listing 6.2: Pseudocode of the function used to fast skip the simulation data without having to reconstruct data linearly each frame.

More information regarding the technical process of the reconstruction of data is given in Section 7.4.1 of the next chapter.

While the reconstruction of data reproduces the trainee performance, the default perspective of an observation using this reconstruction is of the trainee's VR perspective (see Figure 1.1). As the reconstruction is intended to reproduce the trainee performance, the simulation automatically defaults to the camera perspective of the trainee. To to enable assessors to have control of their observation to view the trainee from any perspective, a free camera tool is required.

6.3.2 Free Camera Tool

The requirement collection study conducted with assessors in Section 4.1 identified the requirements of R1, R2, R5 and R6 (see Section 4.2). This required the creation of the free camera tool (see Section 4.3.1) to meet these requirements. The free camera tool enables assessors to have unrestricted control of their perspective during assessment (R1), and provides assessors with the capability of monitoring and observing trainees apply their knowledge to tasked objectives (R2). To meet the requirements of R5 and R6, and extend the capabilities of R2, the free camera tool was designed to be compatible with all other observation, assessment and feedback tools and features embedded within the framework.

Unlike classroom-based assessment, which found that assessors observation of trainee performances could be obscured due to limitations in suitable positions for observation, the framework reconstructs VR trainee performances using 3D co-ordinates within the simulation environment. As such, assessors observation of the trainee can be dynamic, allowing alteration of the assessors observation from any perspective or position within the virtual environment. Furthermore, as observation in VR is after-action, the observation of the trainee is non-invasive, avoiding assessors concerns about causing unnecessary anxiety for trainees (see Section 4.1.1).

To achieve this unrestricted and dynamic observation, the observation component includes a free camera tool that allows assessors to choose where in the simulation environment they want to observe the trainee from. Unlike video forms of assessment, which are restricted to the perspective of the recording (Thurston and Martin, 2011), reconstructing trainee performance in the simulation allow for the trainee performance to be viewed by an assessor from any perspective. The free camera tool provides this functionality, allowing assessors to alter their perspective of observation at anytime. This means that no action or event of the trainee performance is hidden or obscured from the assessor. This implementation can also be used for live observations and assessment, as seen in (Howie and Gilardi, 2020), allowing the assessor to use the free camera tool to focus on the details of the trainee performance without any concerns of causing anxiety for the trainee who will be unaware of the assessors presence (see Section 4.1.1).

Figure 6.4 shows the free camera tool in use, unrestricted by any barrier within the simulation, the camera provides complete control to the assessor to observe the trainee performance from any location within the simulation.

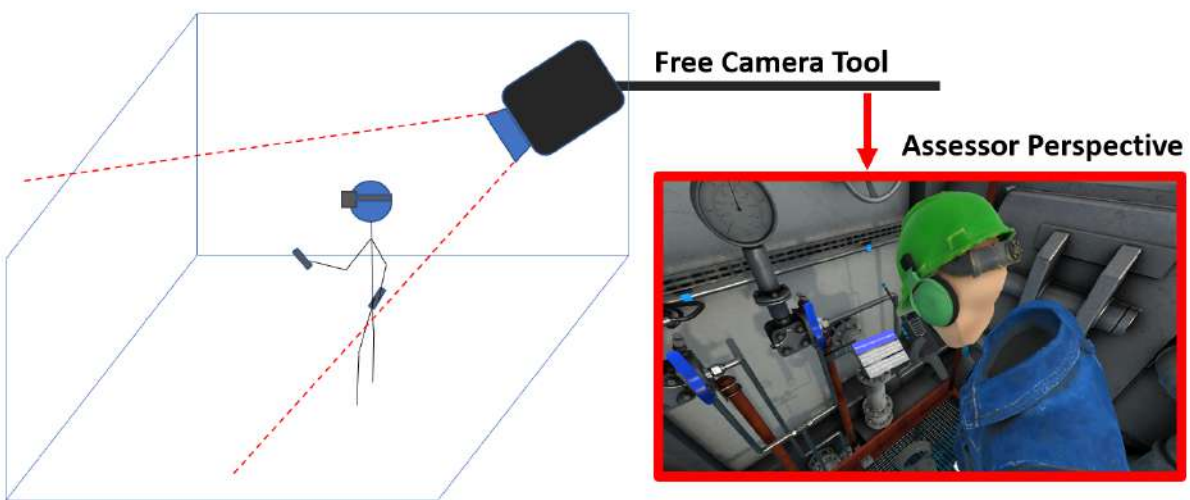


Figure 6.4: A design for the free camera tool (Red Outline) showing the camera perspective viewed by the assessor, along with showing its location within the 3D simulation environment relative to the reconstructed trainee performance.

The free camera tool was designed following the same principles for the default implementation used by the Unity game engine. The assessor controlled the free camera tool using a combination of keyboard input and mouse movement, allowing for six degrees of freedom of movement and rotation.

Due to the use of remote-screen sharing applications during the evaluation, the default free camera implementation standard within Unity was not used. The default implementation from Unity attempted to access the mouse co-ordinates automatically, however as assessors mouse in-

```
--- Camera Controls ----
get input position of the camera using common keyboard commands (W,A, S, D, Q, E, R, F, Space Bar)

newAxisofCamera Equals new Vector3 (-lastAxisofCamera.x - mousePosition.x), - (lastAxisofCamera.y - mousePosition.y), get Axis of
      Mouse ScrollWheel())
lastAxisofCamera Equals mousePosition X and Y co-ordinates.

if rotationButton is pressed
 set speed and sensitivity of camera rotation using newAxisofCamera
else
 make no changes to the camera rotation

update position of the camera
```

Listing 6.3: Pseduocode of the standard free camera control implementation modified to work with remote screen sharing applications (Microsoft Teams, Team Viewer).

put was emulated by Microsoft Teams remote-screen sharing functionality, Unity was unable to automatically detect movement. The alternative implementation was designed around the use of accessing the raw input values from the mouse co-ordinates. By manually monitoring and managing the rotation of the camera using the input mouse co-ordinates and keyboard input each frame, the framework is able to bypass restrictions imposed by screen sharing applications when using remote control functionality. Listing 6.3 shows an extract of the pseduocode used to modify the camera rotation to avoid similar issues when using screen sharing applications.

Because the simulation environment is not restricted by the physical constraints of reality, the observation component is able to provide the assessor with full control of their perspective, avoiding all concerns of intrusiveness or obscured observations, as identified from the background study in Section 4.1. The free camera tool enables assessor to alter their observation position and perspective to gain a better view of the trainees action (see Section 4.3.1), allowing assessors observe the performance of a trainee without any obstructions. However, as assessors may be required to re-watch or re-visit areas of the trainee performance, replay features are needed to enable assessors to rewind and skip sections of the reconstructed trainee performance, as discussed in the next section.

6.3.3 Replay Features

Requirements R1, R2 and R4 (see Section 4.2) were identified from the study conducted with assessors in Section 4.1. Following from the features implemented by the free camera tool (see Section 6.3.2), replay features are required to allow assessors to monitor and observe trainees within simulations for requirement R2, providing assessors with the ability to pause, rewind and skip sections of the trainee performance (see Section 4.3.1). As with the free camera tool, the replay features must be compatible with all other tools and features embedded within the framework to meet the requirements of R1 and R4 (see Section 4.2).

With the dual-recording methodology capturing all data from trainee performances, the observation component is able to use non-linear reconstruction methodology to reproduce the trainee performance. Unlike linear reconstruction methodology which must be reproduced in order from the very start of the data-file, non-linear reconstruction enables reconstruction from anytime in

```
---------- Reproduce Data Function ----------

 If simulationTime is being modified by FastProcessReconstruction OR recordingStatus is NOT EQUAL reconstruction
       return

 If reconstructionMode is Forward
     globalTime add fixed time value
     Reconstruct next available index value for all data
        Forward Rebuild data with new globalTime value
 If reconstructionMode is Rewind
     globalTime remove fixed time value
     Reconstruct previous available index value for all data
        Rewind Rebuild data with new globalTime value

   Clamp globalTime value between 0 and simulationTime
```

Listing 6.4: Pseudocode showing the operation for the framework managing data for reconstruction depending on if the reconstruction is for Forward or Rewind, providing the assessor with flexibility for their observation.

the simulation. Building upon the notion of providing the assessor with complete control of their observation, the replay feature makes use of all data stored within the database (user input, deterministic and non-deterministic data) to alter the time of observation when reconstructing a trainee performance. These feature allow assessors to rewind, fast-forward, slow-motion, speed-up, or skip reconstructed trainee performance while keeping the reconstructed simulation data in sync to enable a consistent reproduction of trainee performances.

These replay features are designed to replicate the features found within video files, such as the ability to reverse, fast-forward or skip sections of a trainee performance (Thurston and Martin, 2011). However, unlike video files which are predefined captures, the observation component must reproduce the trainee performance. For example, if the assessor wanted to rewind the the trainee performance being reconstructed, the replay feature inverts the trainee performance data stored within the database, and reconstructs data in reverse, while keeping the actions and outcomes consistent. Listing 6.4 shows pseudocode for the reconstruction of the simulation based on if the assessor is observing the trainee performance as performed by the trainee (forward), or the observation is operating in reverse, inverting the actions performed by the trainee.

Furthermore, requirements R2 and R4 identified from the study with industrial training assessors (see Section 4.2), showed that assessors were interested in the ability to skip sections of the trainee performance to avoid unnecessary time-wasting when the trainee is reading instructions. As such, the replay feature uses the data stored in the database to enable the use of non-linear reconstruction methodology, which allows consistent reconstruction of the trainee performance from any recorded point of their performance. This means the trainee performance can be consistently reconstructed at any recorded timestamp.

These replay features enable assessors to alter their observation of the trainee performance, providing abilities that are only possible using after-action methodologies (Hanoun and Nahavandi, 2018). However, the purpose of this framework is to enable assessors to conduct authentic assessment and generate contextual feedback of trainee performances. To achieve this, the framework includes embedded tools and features necessary for in-depth assessment of the trainee performance.

However, first it is important to clarify the execution order of data that enables the reconstruction described in this section to operate successfully and consistently.

6.4 Component Execution Order

The framework manages all components in a hierarchical execution order to ensure that data is captured and reproduced at the required times to ensure consistency. To achieve this structure, the framework incorporates a management script, called Virtual Observation Manager, which directs and executes instructions for all components, including the dual-recording components discussed in Chapter 5. This management component is responsible for the framework operation, and depending on if the framework is recording or reconstructing data, dictates all other components. Sub-management script Virtual Observation Tracker handles the capture and recording of data components (lightweight and processed recording), while the other sub-management script Virtual Observation Reconstruction handles all components that focus on reproducing the data (processed recording and observation).

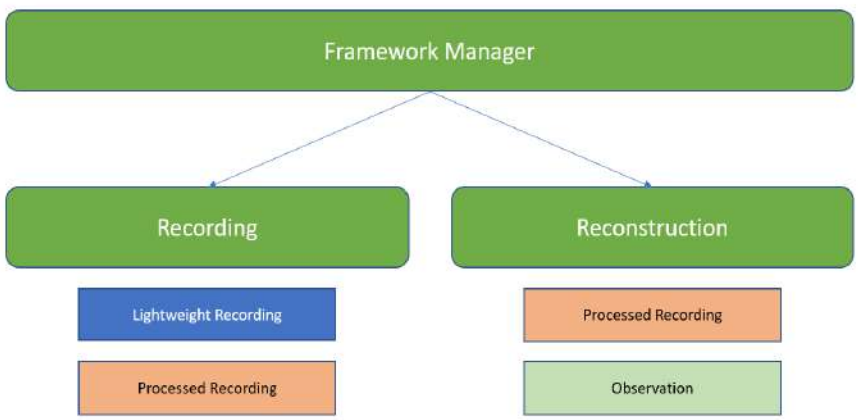


Figure 6.5: The hierarchy of the scripts and their accompanying components from the framework to handle recording and reconstruction. Processed recording features in both scripts, as it is both a recording and reconstruction component.

On awake within the simulation, the framework manager determines the variable conditions defined by the developer or loaded into the scene to determine if operation status of the framework. Depending on the operation status, either the recording components of the framework within *Recording*, or the reconstruction components in *Reconstruction* are activated. Once activated, the sub-management script (*Recording* or *Reconstruction*) determines what components are to complete the requested operation. For example, when recording, the sub-management script determines if the lightweight or processed recording should be activated. This is achieved by determining the variable conditions acquired by the framework manager. The framework manager and sub-management scripts then configure the simulation environment for the tasks required.

For recording, the lightweight and processed recording components operate using the same recording methodology as described in Section 5.3. However, their configuration determines what data is captured.

For lightweight recording, when activated by the framework manager, the component delays the launch of the simulation and automatically begins monitoring user input, and scanning all objects within the current simulation environment to identify non-deterministic variable data. Predefined components such as Unity's rigidbody are automatically configured for recording. As discussed in Section 5.4, lightweight recording can capture entire objects, or can be reduced to capture individual components or variables attached to an object, leaving all remaining non-deterministic data to be automatically captured by the processed recording component. Once lightweight (or processed recording) has configured the environment for recording, the loading of the simulation resumes. At this point, the simulation operates as normal. This delayed launch of the simulation is unnoticeable and lasts less than 1 second.

As discussed in Section 5.5, processed recording configures the environment for recording all non-deterministic data before launching the simulation. To achieve this, the component scans all variables and compares them against those stored in the database to avoid data being overwritten and ensuring that previously recorded data is reconstructed. This requires a mixture of recording and reconstruction. As such, the reconstruction of data is important for both processed recording and observation components.

To ensure data captured and reproduced is consistent, it is important that the order of component execution is timed correctly to avoid discrepancies in data saving or reconstruction operations. For example, if recording of data is incorrectly timed or executed, the data recorded may misrepresent the values of data, leading to syncing issues and invalid values, as previously discussed in Section 2.7.9. To combat this, the framework manages data operations before and after simulation logic.

Figure 6.6 presents the execution order for the components within the framework. This operation is conducted every frame, positioning any reconstruction operations at the start of the update, prior to modifications in game logic. Recording operations are positioned last, after all simulation logic modifications have been conducted to guarantee that the update captured during that frame is accurate and represents that instance within the engine logic.

As shown in Figure 6.6, the reconstruction data components include processed recording and observation, which are focused on overwriting current simulation data to reproduce the same simulation logic. As such, the logic within the simulation is replicated, enabling the reconstruction to be produced in a format viable to capture non-deterministic data or for visual observation by an assessor. When recording data for both lightweight and processed recording, the framework captures data after the simulation logic has been completed. With this logical approach to reconstructing and recording data, the framework is able to produce consistent results for the replication of data that an assessor can use to conduct after-action authentic assessment of reconstructed VR trainee performances.

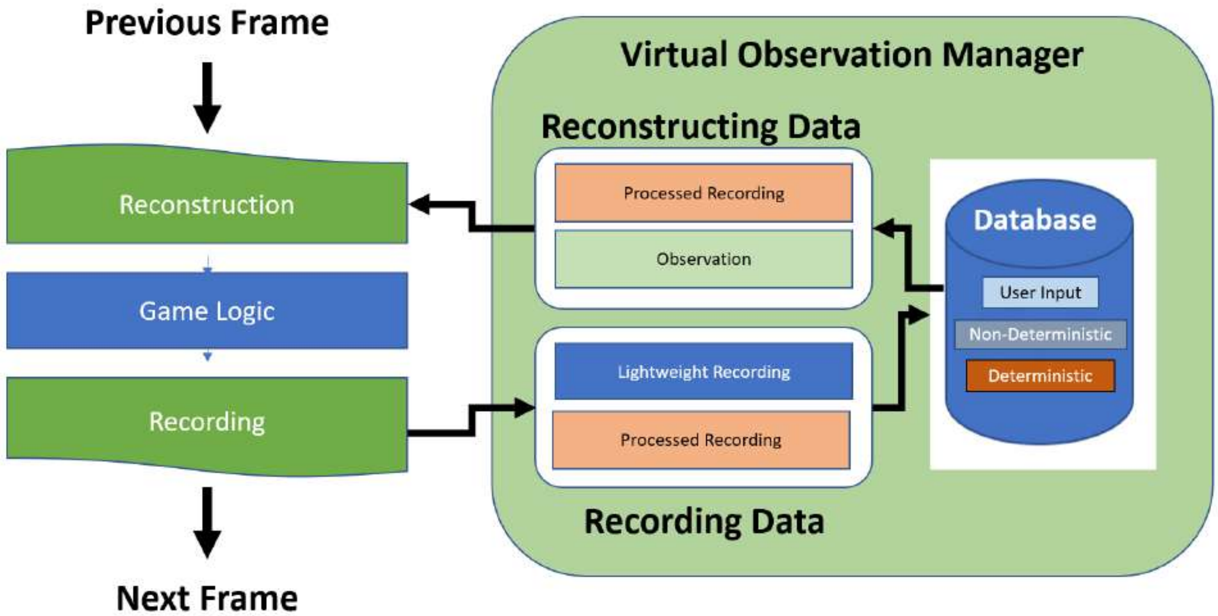


Figure 6.6: The hierarchy of the scripts and their accompanying components from the framework to handle recording and reconstruction.

6.5 Assessment

As previously stated, the purpose of capturing a virtual experiences using the dual-recording methodology described in Chapter 5 is to enable a consistent reconstruction of a trainee performance that is suitable for conducting assessment. To achieve this goal, the observation and presentation of the data must be in a useful and usable format that is necessary to allow assessors to assess the performance of a trainee. The assessment component of the framework expands upon the observation component and includes the assessment tools necessary to meet the requirements R3, R4 and R5 identified in Section 4.2, including summative performance indicators 6.5.1 and the augmented timeline 6.5.2.

An interactive user interface was created for the assessment functionality of the framework. The user interface was based on similar software, mostly inspired by video editing suites (Magix, 2021). The user interface was designed to provide sufficient information to the assessor that presented all necessary information of the trainee performance, and allow assessors to access the assisted assessment tools.

Because a dual-recording methodology was used for the framework, the combined data from both the lightweight and processed recording components stored within the database provide full insight into a trainee performance. As such, this data can be used to generate prior knowledge of the trainee performance, providing assisted assessment tools and features for the assessor that enable in-depth assessment of the trainee actions and their performance completing objectives within the simulation.

6.5.1 Summative Performance Indicators

In section 4.2, it was identified that a requirement of assessors was R3, which required quick identification of the pass or fail objectives outcomes in trainee performances (see Section 4.3.2).

Because most simulations have predefined objectives and tasks to perform, the recorded dataset of the trainee performance will have captured the state(s) of these objective condition variables throughout the trainee performance, providing information about whether objective was successfully completed. Rather than explicitly assigning these variables for recording like most analytical formats (Carruth, 2017; Moore et al., 2020), the data is automatically captured by the dual-recording methodology, which enables trainee performance outcomes to be measured against pass or fail criteria.

For example, Figure 6.7 shows an example of the recorded summative performance indicator data, allowing assessors to determine whether the simulation objectives were achieved based on the outcomes of their task status. The data was dynamically gathered during the reconstruction of data for the observation and assessment components, looking for the variable names that were associated with the task performance of the trainee. This approach could be used to query any data within a simulation, regardless of its perceived importance during development, as discussed in Section 6.2.

During reconstruction of the simulation for observation and assessment, the summative performance indicators tool accesses the database and searches for one or more predefined variable(s) information. This is achieved using a combination of unique gameobject name, component name and variable name to access the variable from the database. The retrieved value of the variable is then compared against a predefined expectation to determine if an objective in the training simulation has been completed. The data is then displayed on the summative interface with passed or failed conditions, with timestamps of the relevant variable value (see Figure 6.7).

The summative performance indicators enable assessor to quickly familiarise themselves with the trainee performance prior to the observation. The summative performance indicators can be provided as a reference to identify when and where a trainee has failed an objective in the simulation, allowing the assessor to pay attention to these areas of the trainees performance, while providing an overview of the objectives completed. The inclusion of this overview information satisfies the requirements identified from the background study with assessors in Section 4.1, which identified summative performance indicators can help assessors perform assessment (see Section 4.2).

Listing 6.5 presents the pseudocode used to create the summative performance indicators values displayed in Figure 6.7, which includes accessing the database to find the necessary information and the function used to format the time based on the saved value retrieved from the database.

As discussed in Section 4.2, assessors expressed the importance of monitoring actions of the trainee throughout the assessment, requiring the framework to have a augmented timeline that

Time: 0:44/13:42	Date: 16/11/21
Objective A	(0:23) PASS
Objective B	(0:58) PASS
Objective C	(1:19) PASS
Objective D	(3:13) PASS
Objective E	(4:20) FAIL
Objective F	(0:00) FAIL

Figure 6.7: A design of the summative performance indicator showing trainee assessment outcomes with Pass or Fail conditions along with their accompanying time marks.

shows the most relevant key-points from trainee performances.

6.5.2 Augmented Timeline

From the study conducted assessors in Section 4.1, it emerged that assessors need to be able to monitor the progress of a trainee during assessment (see Section 4.1). The study identified requirements R1, R2 and R4 (see Section 4.2), which required the creation of the augmented timeline (see Section 4.3.2) to meet the assessors requirement of R4, and partially R1 and R2. As such, the augmented timeline is intended to aid and complement the free camera tool (see Section 6.3.2) and replay features (see Section 6.3.3).

In real-world assessment, familiarisation of the process makes identification of trainee progress easy to acquire, allowing assessors to focus attention on areas of concern where failure is prevalent. However, within VR experiences, it is difficult to adapt that real-world identification of progress to suit changes in assessment approach. This can be partially explained due to the change in context, with the virtual environment replicating a real-world scenario, rather than a classroom-based training configuration. To provide an comprehensive overview of the trainee performance, a augmented timeline was implemented to assist assessors in determining the actions conducted by a trainee against the time of their performance.

Like the summative performance indicators (see Section 6.5.1), the variables presented on the augmented timeline can be altered to suit the needs of the assessment, with any variable from the simulation available in the variable history of the recorded simulation.

The augmented timeline contains a predefined list of variable information that is requested for

```
------Create Menu Text------
    create new instance of the summative performance indicators value text box
    add new instance to list of existing summative performance indicators values
 set object name value of new instance
    set object value using a combination of ConvertTime and FindVariableData
    scale UI text box of the new instance to fit defined interface area

------Convert Time------
get minutes against the floor of globalTime divided by 60
get seconds (remaining) from the remainder time of the minutes calculation
format minutes and seconds into a readable text interface, using minutes and seconds.


------ Find Variable Data ------
for all objects stored in the database
 if object name Equals objectSearchName
     for all components in object
         if component Equals componentSearchName
             for all variables in component
                 if variable Equals variableSearchName
                     return value

if no value found, return an error and alert developer
```

Listing 6.5: Pseduocode used to populate the content within the summative performance indicators UI element.

display on the augmented timeline, depending on the needs of the assessor and objectives within the training simulation. When launching an observation or assessment of trainee performance, the augmented timeline queries the database of recorded trainee data to access the variable history from the predefined list. The retrieved values are then populated on the augmented timeline, creating a new row for each variable, generating the variable history for each timestamped modification, which is scaled to blocks of data that represent the value of the variable. If variable from the database has no modification history, the augmented timeline marks the row as red to alert to the assessor.

The augmented timeline includes three key areas, the variable name, current value of the variable and the complete history of variable information, displayed as blocks of data throughout the augmented timeline (see Figure 6.8 and 6.9). The variable name and current variable value is constantly displayed at the left of the augmented timeline. The related variable information blocks take up the remainder of the augmented timeline screen, and can be zoomed in or out to increase the accuracy of the information against time or display more information on screen at once.

The augmented timeline generates information blocks of data using the database to display data against timestamp modifications. The blocks of data provide information for variables values throughout the trainee performance, with the recording instance timestamps used to identify when and for how long a variable was modified. The augmented timeline can be analysed by the assessor observing the trainee performance, using the data as guidance to observe points of the simulation. For the example in Figure 6.9, the assessor can use the augmented timeline to identify exactly when in the simulation the trainee picks up equipment related to an objective, then skip to that point in the observation using the replay feature functionality (see Section 6.3.3).

Listing 6.6 shows the pseudocode for the augmented timeline identifying and creating objects for display by iterating through the database captured from the recording components (see Section 5.4 and 5.5). Listening 6.7 expands upon this, showing how object data is interpreted by the

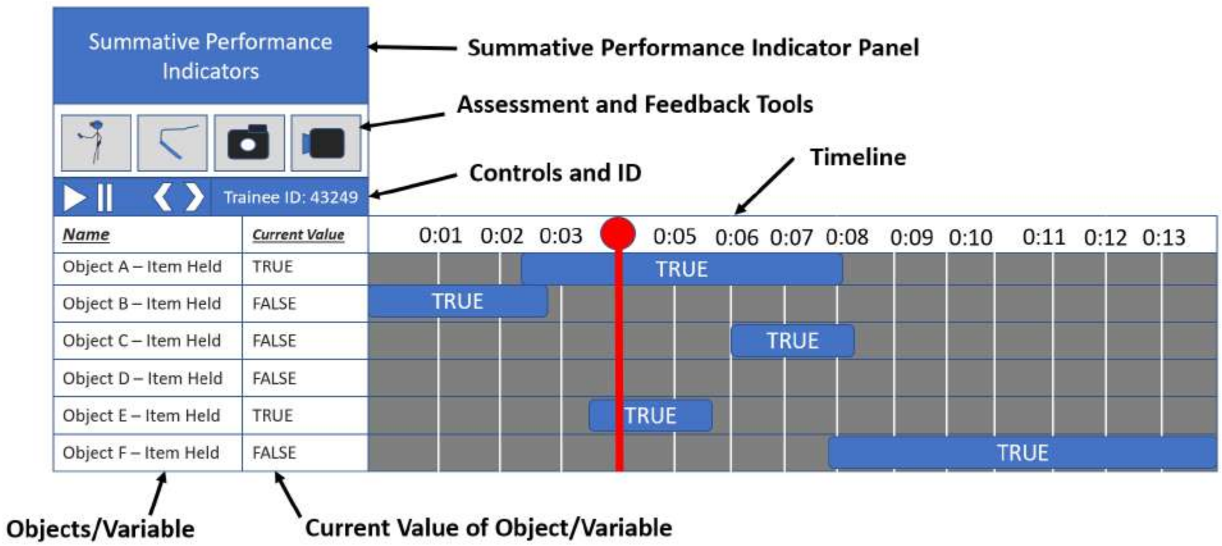


Figure 6.8: The interface layout of the augmented timeline, presenting the play, pause, rewind, forward controls, along with other elements of the interface, including the options to enable the assisted assessment and contextual feedback tools and features. The location of the summative performance indicators interface from Figure 6.7 also shown relative to the augmented timeline.

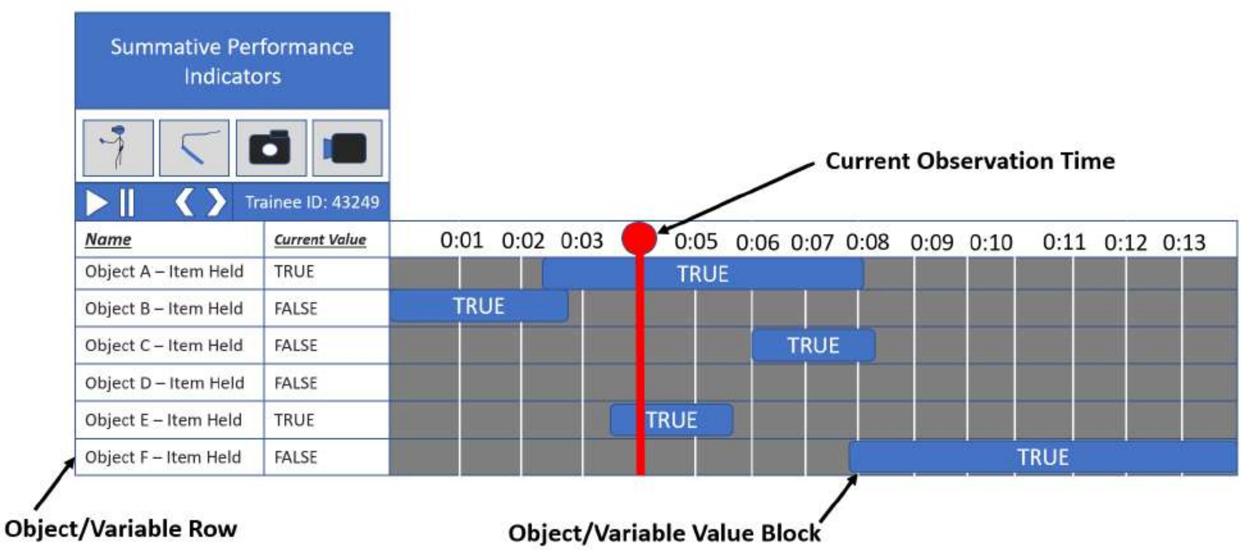


Figure 6.9: The interface layout of the augmented timeline, displaying data for when an individual item was held. This allows assessors quick identification of key-points from the trainee performance, allowing assessments to skip over sections of the performance when a trainee is reading instructions.

augmented timeline to create the block structure presented in Figure 6.8.

```
----- Create Variable Row ------
if variable is not valid to display
    return

```

```
------Identify Variables For Display------
for all object in objectlist
 for all components object
     for for all variables in components
         for all predesignatedVariable names that should be displayed on the augmented timeline
    if variables name matches the current index for predesignatedVariable
                add variables to list for  display on augmented timeline

for all variables for display on augmented timeline
 CreateVariableRow with variable display on augmented timeline
```

Listing 6.6: Pseduocode for identifying how variables were assigned for display on the augmented timeline using the Object Recording implementation.

```
create a new row for the variable
set the active visibility status of the new row as true
add the new row to the list of all rows created for the augmented timeline

add the created row to the currently activated row list
set the last added index to the currently active row list as active
create a value display instance of the variable data
set the text size to fit the size of the display box
set the max font size to the default font size

if the variable data contains only one data count
    define red parameters to the list for the user interface, indicating no change in value

add a new list to populate the values of the current variable data count
name the new list after the current variable name
set the created prefab index to zero
iterate through all counts of the variable data count
```

Listing 6.7: Pseduocode of generating dynamic information for the augmented timeline.

Because the interface for the augmented timeline contains several user interface elements, it was deemed necessary to include a support tool that would inform assessors in real-time, informing them what each element of the interface represented. As such, all elements of the interface included an embedded help tool which notifies the assessor what the content of the UI element represents in-case they are unsure what the element or icon means.

When hovering the mouse cursor over a section of the interface within a single element of UI space for three or more seconds, a pop-up box appears centered at the top right of the cursor. The pop-up box automatically populates its content based on a raycast of the mouse against the UI interface to determine what the user interface element content represents. The identification code acquired from the raycast is then queried to a help database that is designed to support and guide the assessor if they are struggling. The identification code represents predefined support solutions that inform the pop-up box what content it should display to help and inform the assessor. For example, when activated over a block of data displayed on the augmented timeline, the pop-up box displays the name of the variable and current value at the location of the mouse, or informs the assessor the meaning of an icon (Figure 6.10).

Listing 6.8 presents pseduocode used to create the pop-up interface based on the identification code of the UI element for the interface.

The framework extends the capabilities of observation and assessment by including assisted visualisation of the trainee tools, which further assist the assessors to contextualise the trainee performance.

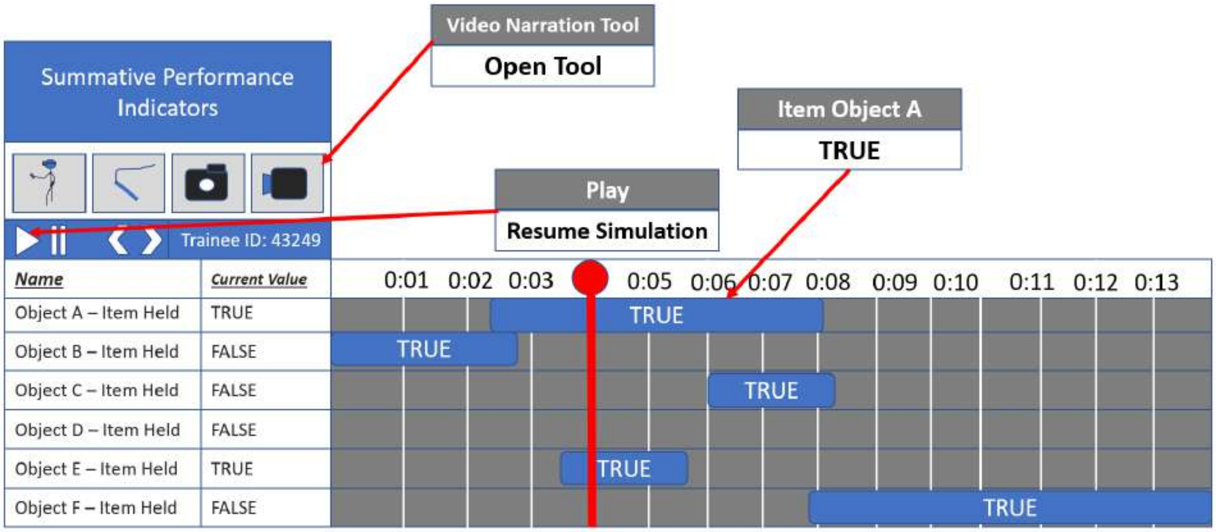


Figure 6.10: Different examples of the pop-up interface being used to guide and support the assessor learn the interface. These pop-ups are activated when the assessor hovers the mouse icon over an UI element for a few seconds.

6.5.3 Assisted Visualisation of the Trainee Tools

The study conducted with assessors in Chapter 4 identified the requirements R1, R2 and R5 (see Section 4.2), Ultimately, these requirements aimed at providing assessors with an unrestricted and unhindered observation of the trainee, which formed the assisted visitation of the trainee tools which complement the free camera tool (see Section 6.3.2) and replay features (see Section 6.3.3). The study included the requirement R5, which sought to replicate trainee performance so that assessors could estimate confidence from trainee performances to identify over or under-confident trainees that may require additional support.

```
----- Create PopUp ------
get identification code
get heading string of pop-up
get content string of pop-up
store reference of raycast activation identification code
scale pop-up interface to fit size of display
position pop-up interface at the location of the mouse cursor + offset to avoid text being obscured

create any other elements required for the interface such as additional rows

set pop-up display as active

----- Update ------

do raycast against UI
if raycast target value NOT EQUAL stored identification code
 update pop-up or disable pop-up based on new position of  mouse location
 break

update pop-up interface location
position pop-up interface at the location of the mouse cursor + offset to avoid text being obscured
```

Listing 6.8: Pseudocode of the pop-up interface being created and updated to display the content of the mouse location against the identification code of the UI element.

Because assessors suggested that posture and body movement was a key sign of estimating the confidence of the trainee (see Sections 4.1 and 4.3.2), assisted visualisation of the trainee tools included a skeleton structure visualisation tool that provides assessors with the ability to estimate the posture of the trainee when completing tasks and objectives. Figure 6.11 shows a prototype of skeleton structure visualisation tool in use within the simulation.

Within VR simulations, avatars are not exact replicas of the user, with feet and legs often removed to improve the acclimatisation of the simulation and reduce potential for any breaks in presence (Jerald, 2016). This implies that avatars within simulations are not accurate representations of the posture of trainees, making assessment of their movement and actions difficult. To address this issue, the assessment component of the framework incorporates a third-party tool (Lang, 2019) to estimate skeleton structure, which visualises the trainee bone positions based on the tracking information of the VR equipment. FinalIK (Lang, 2019) is used for skeleton structure estimation, but any other tool with similar functionalities can be used. Because the skeleton structure is estimated by the tracked VR equipment, there is no guarantee of accuracy, however it provides a reference for assessors to monitor the real-world movement of the trainee while conducting tasks within the simulation.

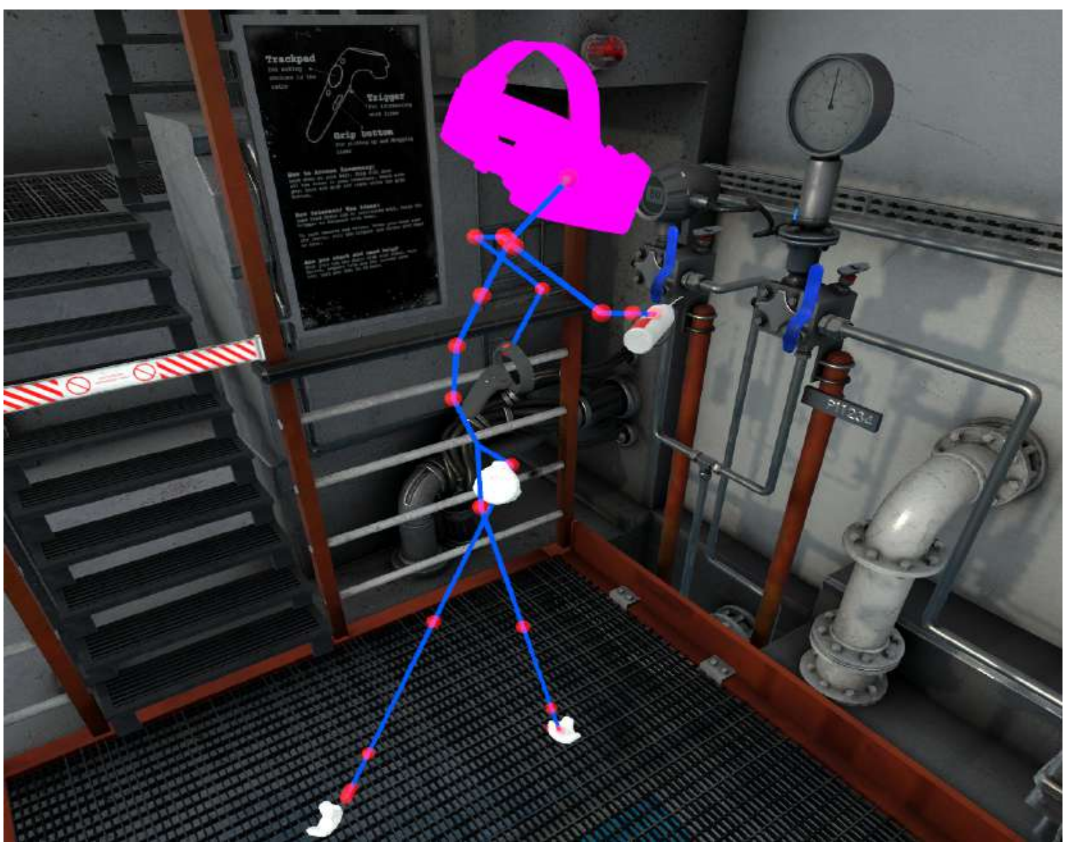


Figure 6.11: Prototype of skeleton structure visualisation in action during the simulation. This visualisation provides an alternative perspective of the trainee that aims to replicate the physical posture of the trainee from the tracked VR equipment (hmd, controllers and trackers).

```
----- Update-----
call generate lines
call draw line

----- Generate Lines -----
for all nodes within body transform
 add or update position relative to tracked VR devices

----- DrawLine -----

set index as 0
foreach transform in childNode
 \\Prevent the drawn lines going beyond the overall structure of the body (eyes, fingers)
 if index is over 2 AND childNode name NOT EQUAL fingerNames AND childNode name NOT EQUAL eyeNames
  Drawline from current position against parent position with current index value
        increase index value
```

Listing 6.9: Pseudocode for the skeleton structure visualisation which updates and modifies the positions of the lines and nodes each frame.

Listing 6.9 presents the pseudocode for the skeleton structure visualisation tool, identifying how lines and nodes are configured each frame to keep them relative to the movement of the tracked VR devices being reconstructed to form the overall skeleton structure.

The requirements R1, R2 and R5 (see Section 4.2), also required the capture of trainee audio to provide additional context to the trainee performance (see Section 4.3.2). As such, the assisted visualisation of the trainee tools included functionality that automatically captured and reproduced the audio from the trainees microphone from the VR equipment. When the simulation is reproduced for assessment, the audio recording of the trainee is replayed in sync with all other elements of the simulation (see Section 6.3.1).

These tools and features enables assessors to hear and see the posture and movement trainee, providing additional audio and visual cues that can be used to estimate confidence levels of trainees, replicating how confidence was measured in classroom-based assessment (see Section 4.1.1).

6.5.4 Live Streaming of Trainee's Sessions

Live streaming of trainee's sessions was presented in (Howie and Gilardi, 2020), demonstrating the effectiveness and portability of the framework for remote live observations and assessment of virtual reality trainees. In (Howie and Gilardi, 2020) it was discussed how an observer can view a VR simulation live during run-time. The implementation allowed the assessor to be present in the simulation environment using either a VR headset or desktop computer.

Live streaming is an important feature of the framework as highlighted in Chapter 4 and addresses the lack of real-time communication between the trainee and assessor during training operations. While in (Howie and Gilardi, 2020) it was showed that it is technically possible to observe assessments live, with data automatically recorded on the assessors computer, further work is required to determine how to best provide assessment tools for the assessor during these live instances since it is not possible to pause, or rewind a live observation in this instance.

Figure 6.12 presents the transmission logic for live streaming data to a server or another computer, allowing the trainee performance to be monitored and observed live in real-time (Howie and

Gilardi, 2020). However, this means that the assessor must observe the simulation in sync with the trainee in a linear format, as the format required for live streaming prevents the processed recording component from operating until all trainee data has been collected and the live session has completed.

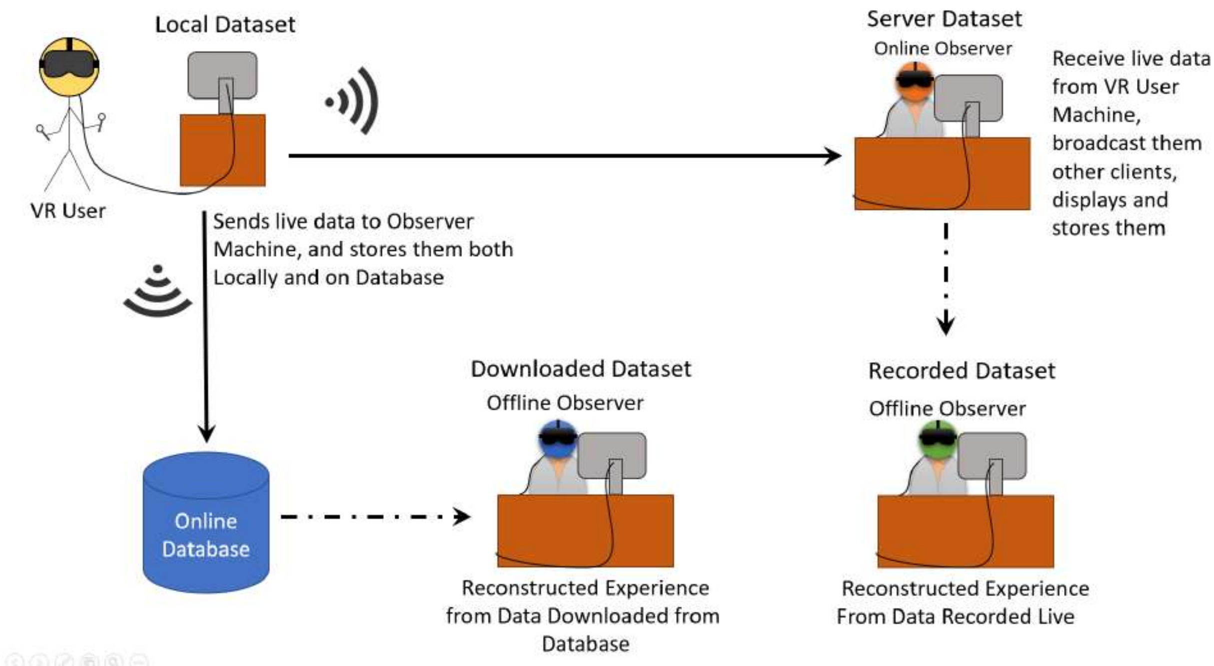


Figure 6.12: As shown in Chapter 1 and presented in (Howie and Gilardi, 2020), the live streaming component enables instant observation of the trainee performances in real-time, along with having the transmitted data locally saved on the assessors computer or sent to a server for after-action review.

While live streaming was not evaluated with assessors in this work, live streaming will play a key-role in connecting trainees and assessors to best replicate classroom-based training, providing the opportunity for instant communication and live teaching and assessment.

In summary, the tools and features provided within the assessment component are designed to assist the assessor by providing an enhanced observation of the trainee performance, with the objective of making the assessment processes easier, useful and usable. However, as it is important that feedback is delivered in a timely manner (Jonsson, 2013), the framework included embedded feedback tools that were designed to replicate the content and context for feedback provided in classroom-based assessments.

6.6 Feedback

The framework and components described in this chapter so far provide the components necessary to record, reconstruct, observe and assess a trainee performance. The last component of the framework is feedback. This component is designed to enable feedback to be conducted within the assessment environment, allowing the framework to streamline the process of conducting authentic assessment and generating contextual feedback. Three additional tools were developed as part of the feedback component: *Video Narration Tool*, *Evidence Capture Tool* and *Drawing Tool*. These tools were designed to meet the requirements identified from the requirement collection study conducted with assessors in Section 4.1.

6.6.1 Video Narration Tool

Requirements R6 and R7 (see Section 4.2) were identified from the study conducted with assessors (see Section 4.1), highlighting the importance of providing feedback that is contextual and supportive for the trainee, which allows them identify and understand what went wrong and correct their gaps in knowledge.

To provide contextual feedback of a trainee performance, several solutions were possible, as discussed in Section 2.8. However, since written feedback requires additional time and effort on behalf of the assessor, video feedback was deemed best suited, as it provided visual contextual to the feedback (see Section 4.3.3).

The video feedback tool captures video recordings from the perspective of the assessor observing the trainee performance, capturing exactly what the assessor is looking at, including any movement or changes in perspective from the virtual free camera tool (see Section 6.3). As such, the video narration tool was designed to be compatible with the free camera tool (see Section 6.3.2) and video drawing tool (see Section 6.6.3). To enhance the content of the feedback for the video tool, audio recording from the assessors microphone were captured and embedded into the generated video, enabling incorporation of automatic voice-over commentary. The technical aspects for recording the video was handled by Video Capture Pro (EveReal, 2021). This plugin enabled the video and audio recordings captured by the assessor to be exported and accessed outside of the Unity engine environment (EveReal, 2021).

The video narration tool provides both context and reference to the feedback generated by the assessor. The output video was designed to replicate as close as possible, the feedback experience of classroom-based assessment. The video narration tool generates a standard mp4 video file. This makes the feedback portable and easily accessible by the trainee, which is only constrained by the storage space required to save the video file for archiving.

For documenting and archiving failures, an evidence capture tool was designed to capture screenshots during observation and assessment of a trainee performance.

6.6.2 Evidence Capture Tool

In Section 4.2 the requirement R7 was identified, which required the creation of a evidence capture tool that could archive evidence of training failures (see Section 4.3.3). The evidence capture tool allows assessors to archive evidence of trainee failures, in-case a trainee disputed their assessment outcome. While disputing assessment outcomes is considered rare by assessors (see Section 4.1), it represents a problem that may be more common in virtual forms of assessment, as trainees may feel inclined to dispute any assessment of their performance on the assumption the trainee would be unable to dispute their claims. While the framework enables the entire trainee performance data-set to be archived itself (see Section 6.2), or could be video recorded using video narration tool discussed in Section 6.6.1, capturing a picture of a failure can be useful for quicker access, portability and low-storage space.

The evidence capture tool captures a single image of the trainee session from the assessors perspective, outputting an image. However, both the video narration tool and evidence capture tools lack the ability to guide and focus the observation gaze of the trainee when they receive the feedback. To address this limitation a drawing tool was designed.

6.6.3 Drawing Tool

A crucial feature of feedback identified from the study presented in Section 4.1 was feedback presentation. Results from this study identified requirements R6 and R8 (see Section 4.2), which showed that assessors highlight concerns during feedback by guiding a trainees attention to desired areas of interest using gestures and physical demonstrations to provide contextual and targeted feedback. When a trainee struggled to understand an objective the assessor frequently points to each individual element of the process, followed by verbal explanation, demonstrating the steps to a trainee. Likewise, during trainee practice work, assessors highlighted and corrected mistakes by guiding trainees attention to mistakes by pointing and explaining the trainees error. The gesture is intended to focus the gaze of the trainee to a specific area of the work environment.

The use of gestures is considered a method of grounding speech to a physical environment, clarifying meanings from verbal messaging to physical constraints in the environment, informing the trainee what to focus their attention on (Valenzeno, Alibali, and Klatzky, 2003). In virtual environments, it may not be possible for the verbal and visual feedback to be correctly grounded to a specific area of a simulation alone.

To replicate the use of gestures and to provide a tool that can dictate the focus of the trainee, a drawing tool was implemented into the framework (see Section 4.3.3). The drawing tool allowed the assessor to ‘draw on top of’ the screen, so that lines and drawn text could be embedded in the assessor perspective in both the video narration tool and evidence capture tool, overlaying the lines into the video and feedback pictures captured. While this tool is not a like-for-like replacement for real-world gestures, it is an accessible implementation that an assessor can use to direct attention and highlight objectives missed in an overall effort to provide feedback that is beneficial for the trainee. The drawing tool was designed to be compatible with the free camera tool, video narration tools and evidence capture tools.

Figure 6.13 depicts the output generated by the video narration tool that combined the use of the drawing tool to create a contextual and targeted feedback for the trainee.

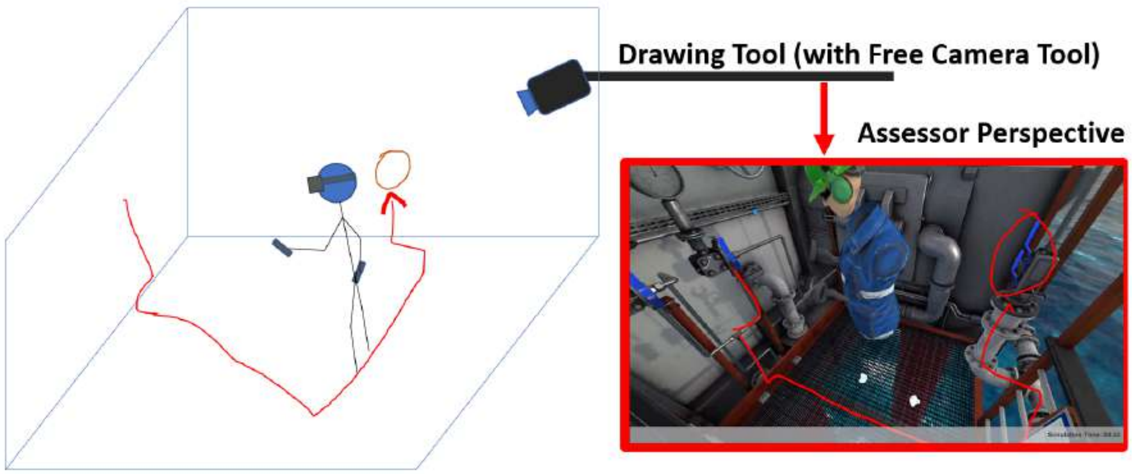


Figure 6.13: A presentation for how the drawing tool and free camera tool is used to generate feedback. The output from the video narration tool and drawing tool is in the red outline, providing contextual and targeted feedback that combines audio, visual and gesture cues.

Listing 6.10 shows an extract of pseudocode that was used to implement the drawing tool functionality. As it was important that the drawing tool did not overlap with the UI elements, the drawing tool performs several checks to confirm that the drawing location is valid within the world co-ordinates of the simulation environment. Unity's line renderer functionality is used to render the lines within the co-ordinates relative to the screen position of the free camera tool (see Section 6.3.2). Rendering the drawn lines relative to the screen position was considered the most suitable approach, however, this means if the camera position is moved, the line drawings are lost as their positioning is relative to the perspective of the camera. An alternative solution is to draw the lines relative to objects within the simulation, which the framework included when using room-scale VR equipment to perform the observation and assessment. This alternative implementation operated as a floating pen in the 3D space, allowing drawn lines to be drawn relative to the objects within the simulation environment, rather than to the perspective of the observer. However, this 3d space functionality is not suitable for the desktop implementations, as the medium is difficult to adapt and control when not using immersive equipment that allows for natural interaction within the environment using hands and physical motions.

```
----- Late Update -----

if drawingTool is NOT active
 return

if observationMode is Desktop (different functionality is required to handle drawing using VR equipment)
 offset distance from the screen in front of the camera position
    perform check to see if the mouse is over a UI element
    if the drawingButton is DOWN
     set mouseDrawingPosition as current Input Mouse position
        use offset distance from screen
        perform a screen to world matrix calculation based on the mouseDrawingPosition
        StartDrawingLine at matrix calculation
        set paintingIsActive as true

else if drawingButton is UP
 setPaintingIsActive as false

if paintingIsActive EQUALS true AND mouse is NOT over a UI element
 perform drawing operation at current mouse position using screen to world matrix calculation

----- Drawing Operation -----
if worldMousePosition is MORE THAN overlapDistance OR this drawing operation is at index 0
 Add new index to line renderer tool
    Set new index to worldMousePosition
    Update line renderer tool
```

Listing 6.10: An extract of pseudocode showing how the drawing tool is implemented to avoid drawing over UI elements and preventing overlapping lines to be drawn, while maintaining a smooth drawing operation.

6.7 Chapter Summary

The purpose of this chapter was to present the remaining components of observation, assessment and feedback and their accompanying tools and features as part of the framework that enables assessors to conduct authentic assessment and generate feedback for room-scale virtual reality training performances.

Following the results from the study with assessors (RQ3,and RQ4), presented in Chapter 4, the remaining components of framework answer RQ5 and meeting the requirements identified by the requirement collection study with assessors (see Section 4.3).

The five core components of the framework are: lightweight recording (see Section 5.4), processed recording (see Section 5.5), observation (see Section 6.3), assessment (see Section 6.5) and feedback (see Section 6.6), with the latter three presented in this chapter. When combined, the framework captures all data within a simulation for trainee performances, reconstructs it for direct visual observation to enable assessors to conduct authentic assessment, and generate contextual feedback of room-scale VR training simulations.

To answer RQ6 and RQ7, the framework and its embedded components, tool and features must be evaluated with assessors to determine its usefulness and usability. To conduct this evaluation, a technical implementation of the framework was created, this development process is presented in Chapter 7, demonstrating algorithms and design principles of the most technically challenging components, tools, and features, which were developed for the Unity Game Engine.

Chapter 7

Technical Implementation of the Framework

7.1 Introduction

To evaluate the framework with assessors, an implementation was required. This chapter will discuss the core technical elements of the frameworks components, tool and features that were presented in Chapter 5 and 6, forming a technical implementation of the framework. As such, this chapter presents the technically challenging code and algorithms which enables the architecture of the framework and its embedded components, tools, and features to operate seamlessly together.

In this chapter, the background implementation of the framework in the Unity Game Engine is provided, followed by the technical implementation of the dual-recording methodology, including the lightweight and processed recording components. The reconstruction methodology follows, presenting how the framework architecture enables consistent reconstructions of trainee performance data, using a dynamic library system to convert all saved data back to the their variable type. Next, technical details behind the assessment, observation and feedback components are provided, showing how code used to implement some of the tools and functionalities that were implemented into the framework.

The framework detailed in this work shares similarities to previously published iterations (Howie and Gilardi, 2019; Howie and Gilardi, 2020). These previous iterations acted as proof of concepts to determine if the virtual approach to recording and reconstruction was feasible. However, as discussed, the previous published implementations had limited in flexibility and were constrained due to the approach taken for capturing and reconstructing the data (see Section 5.1). This led to the creation and development of the dual-recording methodology (see Chapter 5).

7.2 Script Execution Order

As discussed in Chapter 5, the dual-recording methodology categorises data into: *user input*, *non-deterministic*, and *deterministic*. The technical implementation of the framework operates on the knowledge that the all data within a simulation fits into these three categories. The framework was developed for the Unity Game Engine. While the implementation of components and the overall methodology of the framework described in this chapter are specific to the concepts for the Unity Game Engine, the core components are universally applicable for other game engines or development environments. Unity's order of execution for event functions is shown in Figure 7.1.

To enable consistent recording and reconstruction, the scripts that handle these processes are set to custom execution ordering, which dictates what order the Unity engine executes them. Scripts labeled with negative numbers are executed first, with the lowest negative value script being the first script to be executed each frame. Scripts labeled with 0 are executed in normal execution order, with no delay. Scripts labeled with positive numbers are delayed executions, which take place after all other scripts, with the larger value scripts being the last scripts to be executed each frame.

The script execution of the framework can be viewed in Figure 7.2. Three scripts were utilised to execute a consistent recording and reconstruction operation: *VirtualObservationManager*, *VirtualObservationTrackerReconstruction*, and *VirtualObservationTracker*. These scripts correlate to the lightweight recording, processed recording and observation components, as described in Chapters 5 and 6.

Expanding upon the design overview presented in Section 6.4, the first script to be executed during a frame is the VirtualObservationManager (VOManager), which handles all components and systems for the framework, setting up recording and reconstruction operations. As discussed in Section 6.4, all other scripts (functions) link back to VOManager at some-point during their operation to access data regarding the recording/reconstruction time and access other components of the framework. VOManager is executed first as it initialises all other conditions that make recording and reconstruction possible, such as assigning the recording details and configuring what data is to be saved/reconstructed.

VirtualObservationTrackerReconstruction (VOReconstruction) is next script to execute. VOReconstruction handles all operations for reconstructing data, allowing transformation and logic data to be separated to ensure the reconstruction matches the execution of data modification captured during recording (see Section 6.3.1). This means that where necessary, transformation data (position, rotation, scale) is set before physics collision detection takes place, with logic data updating afterwords.

VirtualObservationTracker (VOTracker) provides the recording components of the framework, and includes all functionality for monitoring and saving variable data (see Sections 5.4 and 5.5). By recording data after all other scripts and executions have taken place, the recording operation is able to capture the final value data for that individual frame. If the default time was kept and

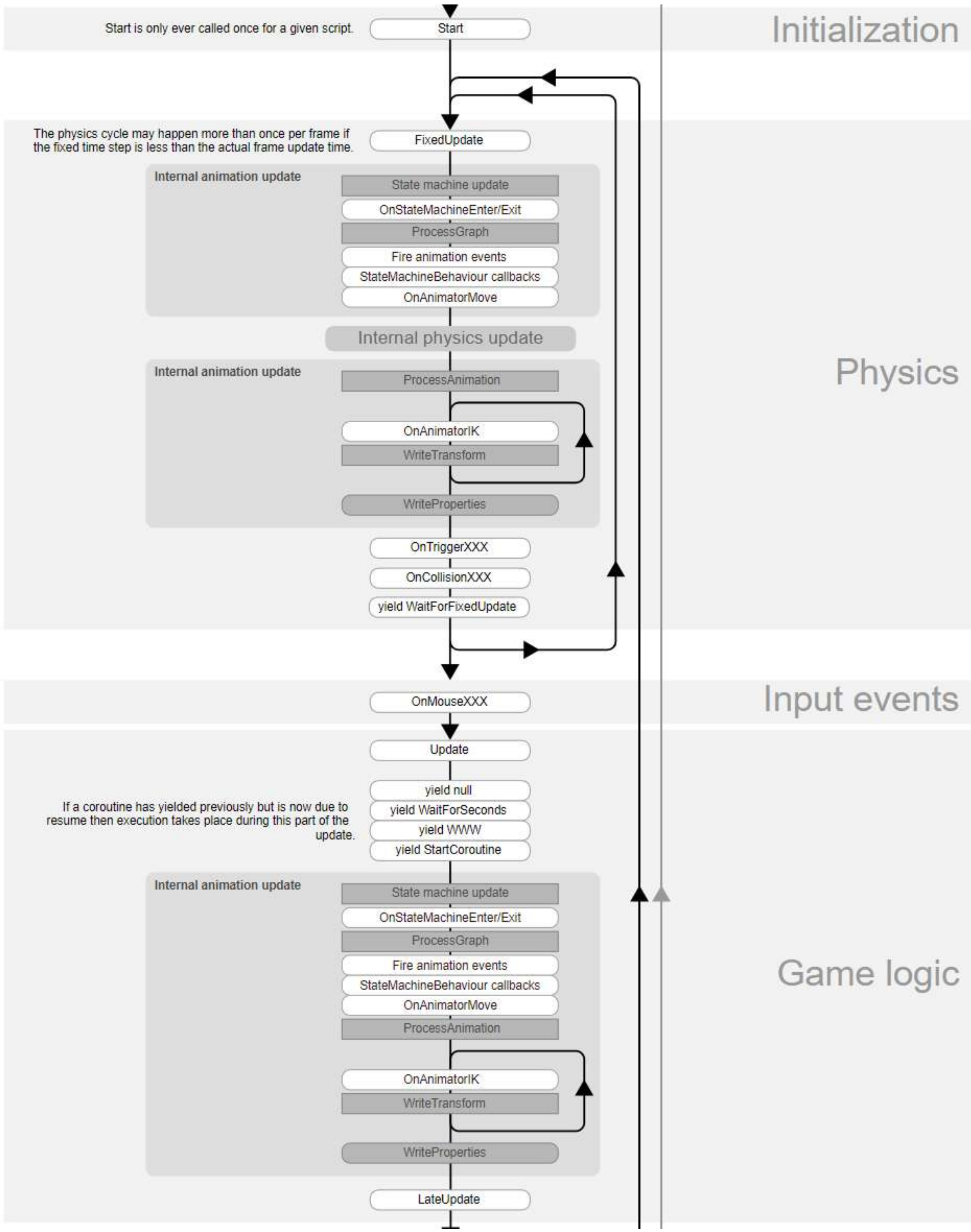


Figure 7.1: Unity's order of execution for event functions (Unity, 2021b) (Image published with Unity Technologies permission).

VOTracker captured data prior to the end of the frame, the recording could miss out on data that is modified by simulation scripts executed after the recording operation has been complete. As a consequence, it is important that the script that handles the recording components is executed as

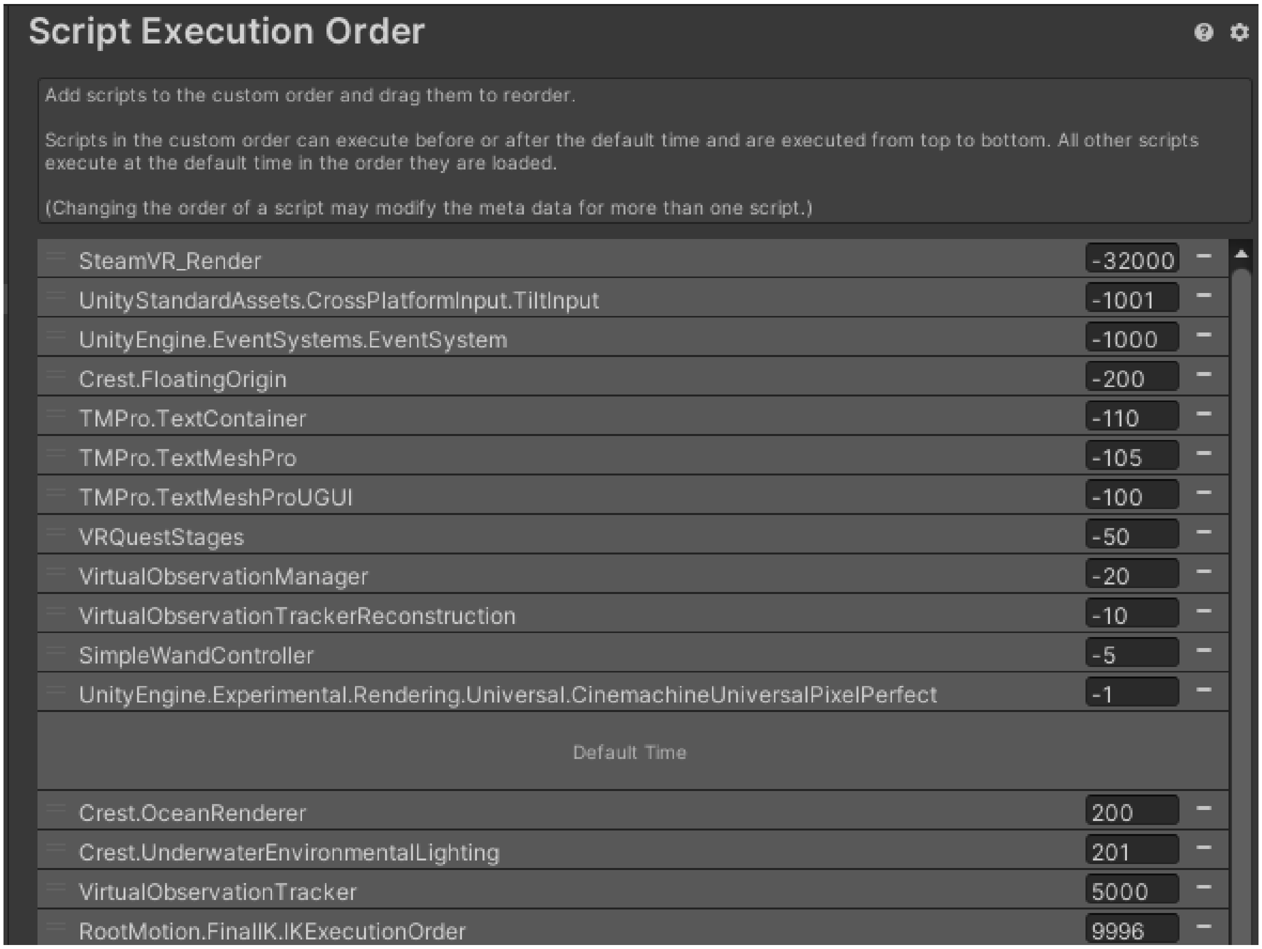


Figure 7.2: The script execution order of the framework scripts as previously shown in Figure 6.5. VOManager manages the recording and reconstruction operations, along with all components of the framework. VOTrackerReconstruction handles all operations that require reconstruction of data, and VOTracker handles all recording operations for both the lightweight and processed recording components.

the last operation for any logic function within a recording operation.

These scripts are used to embed the components of the framework into an execution order that enables a logical capture of data, with reliable reconstruction to ensure the simulation is consistently reproduced (see Section 6.4).

7.3 Lightweight Recording

As documented in Section 5.3, lightweight recording captures user input, and non-deterministic data. To ensure the user input can be both recorded and reconstructed, an input configuration is required to capture all necessary input from the trainee, such as the movement and controller input of the VR equipment. This is conducted by generating an input configuration into the framework.

```
public bool Input(GameObject controllerDevice, List<InputDevice> controller, VRActorObject _heldObject, GameObject _highlightObject
    , GameObject _gripObject, Vector3 _controllerVelocity, Vector3 _controllerAngularVelocity)
{
    InputValueUpdates(controllerDevice);
    bool input_changed = false;
    if (controller.Count >= 1) //identifying if a controller is active
    {
        controller[0].TryGetFeatureValue(CommonUsages.trigger, out _XRtriggerValue);
        controller[0].TryGetFeatureValue(CommonUsages.grip, out _XRgripValue);
        controller[0].TryGetFeatureValue(CommonUsages.primary2DAxis, out _XRprimary2DAxisValue);
    }

    touchPadAxis = _XRprimary2DAxisValue;
    secondaryAxis = _XRsecondary2DAxisValue;
    triggerValue = _XRtriggerValue;
    SetInputStatus(ref triggerButton, t_triggerButton, "triggerButton", _XRtriggerButtonValue, ref heldTime_triggerButton);

    //Comparing previous values to current values determines if any input has been detected.
    input_changed = ((triggerButton != t_triggerButton) || (gripButton != t_gripButton) || (menuBttn != t_menuBttn);
    return input_changed;
}
```

Listing 7.1: Input data configuration. Each frame the input is configured to detect all changes in input. This can be extended as necessary to include/remove additional input functionality. Any additional input added/removed here is automatically detected by the monitoring and saving process, which can handle the data dynamically.

7.3.1 Input Configuration

The input configuration is the core of the structure of what makes this work possible. A custom InputManager was created which consisted of every possible input interaction from a VR controller device, using the accessible data from the development SDK as a basis for modifiable input values. These values were named to correlate to their respective button prompt on the controller, making identification and implementation easy to handle. Each frame, the gameobject representing the controller in the virtual space was passed into the attached InputManager which clarified the changes in input by comparing the current frame status of each button against the previous frame status. If a change was detected, the button state would change to reflect the new state of the controller, button input types were classified into custom states for Idle, Down, Held and Up, which provided sufficient information regarding how a system would interpret the input actions. By encapsulating all user input into an InputManger, all logic for the trainee operating the simulation must go through this script execution, making it possible for the data to be reliably recorded and reconstructed, with the simulation responding consistently to the actions, regardless of if they are from live input from a trainee at run-time, or reconstructed.

Listing 7.1 shows key parts of code used to achieve the input recording functionality. At each frame, the ‘input’ (Listing 7.1) function is executed as part of the normal execution of the Unity engine. The input accesses the SDK of the VR equipment to determine the values of input data, and then assigns them to a stored value which encapsulates all user interactions. All variables captured are then compared for their existing frame, returning a true or false condition if user input was modified. This enables the framework to be made aware that user input has changed which is only necessary for live-streaming the data, but is useful for debugging and analysis.

The framework considers the InputManager as the pass through for all user modified input from external interaction devices, such as controllers and eye-tracking equipment. Therefore it is necessary to record the InputManager data in sync with other non-deterministic data to ensure the recording and reconstruction of the trainee performance is consistent.

7.3.2 Recording Data

As discussed in Section 6.4 and 7.2, the execution cycle of the Unity engine dictates the order each script will be executed during each update cycle. Because recording is designed to capture all data modified in each frame, recording data operations are conducted last in the execution order. This ensures that data is recorded after logic modifications have been completed for that frame.

The recording operations are conducted during Fixed Update and Late Update sections of the rendering cycle of the Unity engine. User input data (see Section 7.3.1) is captured during Fixed Update, with all remaining non-deterministic data captured during Late Update. Lightweight recording splits the recording operations as user input is time sensitive. As such, user input is captured after the input has been registered within the game engine update frame (see Section 7.3.1 and Listing 7.1). The recording of non-deterministic data is delayed until the end of the frame as part of Late Update. This ensure that non-deterministic data is captured only after all possible modifications have been conducted. For example, if user input resulted in the modification of a non-deterministic object, it is important that this logic operation is conducted prior to the non-deterministic data values being recorded. This same operation is used for processed recordings, which again waits until the end of the frame during Late Update to record data, allowing for logic modifications to be completed.

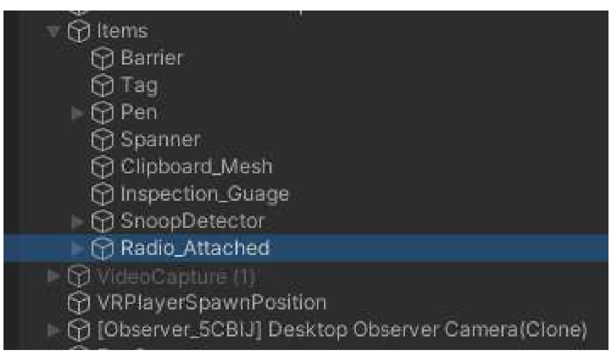
Furthermore, recording operations are split into: *user input* and *object recording*, which separate user input from interaction devices and internal object and simulation logic data from the game engine. The implementation of object recording follows the data hierarchy of data discussed in Section 5.3.4, with object recording designed to capture all information from individual objects within the training simulation, which includes both non-deterministic and deterministic data.

As described in Section 7.3.1, *user input* data includes all variables input by the trainee acquired from interaction devices, including controller input, motion, and audio, but can be extended to suit the requirements of the simulation, to include additional data like eye-tracking or other kinematic data, such as face or body tracking. The user input is captured during Fixed Update, prioritising recording in sync with the user input, enabling easy incorporation into the live-streaming component (Howie and Gilardi, 2020) for future work.

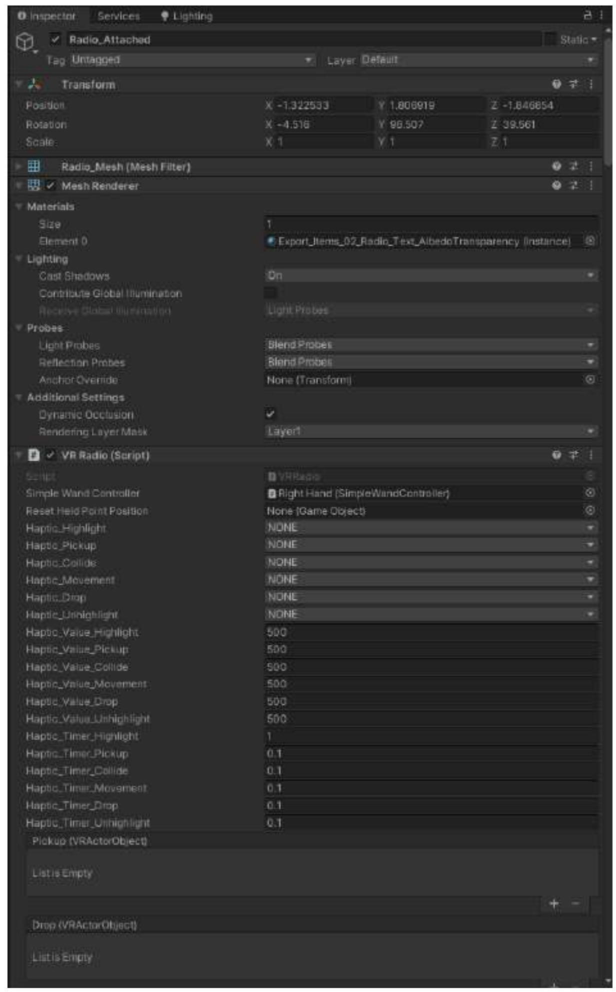
Object recording is the main recording operation that captures data for objects within the simulation environment. During lightweight recording, object recording captures non-deterministic object variable data, such as physics or AI driven components. Processed recording conducts the same recording operation, using object recording to capture all deterministic data, filling in the gaps for the recording hierarchy (see Section 5.3.4).

Both recording systems operate using the same core logic, but use different structures. Object recording replicates the hierarchy used for objects in the unity game engine (see Section 5.3.4), whereas, user input is encapsulated into its own section of the recording hierarchy. This means the user input, such as movement and controller input are structured to form the user input modifications. All objects within the simulation environment are structured replicate the recording hierarchy.

Object recording replicates the representation used by the Unity engine, structuring the hierarchy of data as: an individual GameObject, Components attached, and the variables attached to each component (see Section 5.3.4). This structure represents the overall hierarchy of game data during a simulation, enabling consistent recording and reconstruction (Figure 7.3) of data at an individual object, component or variable level.



(a) Extract of the object hierarchy in Unity.



(b) Extract of the radio (Radio Attached) object component structure for the Radio, showing the components and variables embedded within each component (see Figure 6.2).

Figure 7.3: (Left) Extract of the object hierarchy in Unity, (Right) Extract of the radio (Radio Attached) object component structure for the Radio, showing the components and variables embedded within each component, as seen in Figure 6.2.

```
public void TrackData(GameObject _gameObject, float time)
{
    if (_gameObject.transform.parent == null)
    {
        parent.TrackData(time, "null");
    } else {
        parent.TrackData(time, _gameObject.transform.parent.name);
    }

    enabled.TrackData(time, _gameObject.activeInHierarchy);
    position.TrackData(time, _gameObject.transform.position);
    rotation.TrackData(time, _gameObject.transform.rotation);
    scale.TrackData(time, _gameObject.transform.localScale);

    if (SystemDataInstance == null) return;

    for (int i=0; i<SystemDataInstance.Count; i++)
    {
        SystemDataInstance[i].SaveData(_trackedGameObject, time);
    }
}
```

Listing 7.2: Code used to save gameobject, component and variable data.

For both the lightweight and processed recording components, the object recording function captures data by iteratively going through each object, their components, and then the variables determining if their value has been modified, and if so, instating a recording instance (see Section 5.4.1). Listing 7.2 shows the monitoring and recording operations for *Object Recording*.

Listing 7.2 operates by accepting a object from the simulation, then tracking its transform data, including parent, status, position, rotation and scale. Following that, the function determines if the data includes any child objects (sub-elements) of the object. If so, the data then iterates through them, storing them under the hierarchy of the main object passed into the function. At each function check, the current time of the simulation is passed in along with the current relevant value of the objects or data element. Using the position data as an example, the current position of the object is compared against the previously recorded value. This determines if the value has been modified.

Listing 7.2 shows the iterative loop used to save each variable attached to a component or object, passing them into the monitoring and saving function that replicates the hierarchy. If the value of the variable is different from the previously stored value, the data would be further transferred to the *SaveDataInstance* function displayed in Listing 7.3. The *SaveDataInstance* function then accesses the raw value of the data using system reflection, accessing the variable as type *object*. This *object* data is then converted to string and saved with a timestamp of the data element used. The function supports both Property and Field types.

Because the system uses system reflection to store data (see Section 5.3.2), the monitoring and recording of data can be conducted dynamically. This means regardless of what variables are used within a simulation the functions demonstrated are robust enough to be able to access the previous and current values, then determine if saving is necessary. As such, VOTracker monitors and saves variable data every Late Update call for the Unity engine, creating a detailed and in sync log of all variable modifications, saving them to the database (see Section 6.2).

The recording and storage of user input is conducted slightly differently, but follows the same principles as above, with user input data stored to reference an individual unique variable within a

```
public void SaveDataInstance(object _object, float time)
{
    for (int i = 0; i < systemDataInstance.Count; i++)
    {
        object varData = null;
        if (systemDataInstance[i].dataInstanceType == SystemDataInstance.DataInstanceType.Property)
        {
            varData = _object.GetType().GetProperty(systemDataInstance[i].variableName, System.Reflection.BindingFlags.Instance | System.Reflection.BindingFlags.Public).GetValue(_object, null);
        }
        else if (systemDataInstance[i].dataInstanceType == SystemDataInstance.DataInstanceType.Field)
        {
            varData = _object.GetType().GetField(systemDataInstance[i].variableName, System.Reflection.BindingFlags.Instance | System.Reflection.BindingFlags.Public).GetValue(_object);
        }
        systemDataInstance[i].Save(varData.ToString(), time);
    }
}
```

Listing 7.3: Code used to save individual variable data.

simulation. The process is similar to the above functionality, the only notable difference is that the recording of user input operates during Fixed Update, rather than Late Update, and the structure variable data for user input is saved outside of the hierarchical structure used for object recording. This means user input is identified as its own object within the recording hierarchy at the top of the database, allowing quick access to it. This approach was taken as input from the user may be derived from multiple trainees, therefore, by structuring user input separately from all other objects within the simulation environment, it is easier to manage during development and allows for separation of the data that can be useful for debugging and testing of simulations, as demonstrated in (Howie and Gilardi, 2019).

All data captured was saved to Json files (but XML is also supported) to the local computer running the simulation. However, as demonstrated in (Howie and Gilardi, 2020), remote transmission of data to a server or external computer is possible, allowing all clients to capture a localised version of the simulation if necessary. Listing 7.4 shows an example of the Json file architecture used to store data. The example is taken from data captured using the function from Listing 7.2. Listing 7.4 shows how the architecture of the recording is used to keep track of modifications of game data, displaying each object saved has a list of values and times, which determine the history of modifications, time of modification, and what each modification was.

Listing 7.5 shows an extract of code used to create the saved database hierarchy presented in Figure 7.4 and discussed in Section 5.3.4, which structures data based on the logic and hierarchy used by the Unity engine.

In the current implementation, the framework identifies variables by indexing against their object name. Because all object names were unique in the simulation used for evaluation, this was not an issue. However, if two objects of the same name are present in the same simulation at once, only one instance of the object would be recorded. In future implementations, each object stored in the object recording hierarchy should be unique defined with an identifier key.

A limitation of this work is that objects which are dynamically created at run-time during a simulation are not recorded by the framework, since the architecture only detects objects initially present when the simulation starts. An example of how this limitation effects recording can be

```
        {
            "gameobjectIdentifier": "Spanner",
            "gameobjectUniqueIdentifier": "Spanner",
            "_enabled": {
                "variableValue": [],  //Variable Values populated here
                "variableTimer": [], //Times of variable value change here
                "index": 0
            },
            "_position": {
                "variableValue": [], //Variable Values populated here
                "variableTimer": [], //Times of variable value change here
                "timeVal": 0.0,
                "index": 0
            },
            "_rotation": {
                "variableValue": [],  //Variable Values populated here
                "variableTimer": [], //Times of variable value change here
            },
            "_scale": {
                "variableValue": [], //Variable Values populated here
                "variableTimer": [],/ /Times of variable value change here
                "timeVal": 0.0,
                "index": 0
            },
    "systemDataComponentInstance": [ //Components attached to the gameobject.
                {
                    "systemDataInstance": [ //Variable attached to the first component of the gameobject
                        {
                            "dataInstanceType": 1,
                            "variableName": "MyCustomClass",
                            "variableType": "MyCustomClass",
                            "variableValue": [],
                            "variableTimer": [],
                            "ExtendedVariables": [],
                            "_isList": false,
                            "listIndexPosition": -1,
                            "index": 0
                        }
                }
            }
```

Listing 7.4: Extract from saved recording, displaying the gameobject data (enabled status, position, rotation and scale), followed the components and variable data This structure replicates the Unity recording hierarchy (see Figure 5.3.4).

```
public void SaveGameObject(GameObject _gameObject)
{
    gameobjectIdentifier = _gameObject.name;
    gameobjectUniqueIdentifier = _gameObject.name;
    Component[] _components = _gameObject.GetComponents<Component>();
    for (int i = 0; i < _components.Length; i++)
        SaveComponents(_components);
}

public void SaveComponents(Component[] _components)
{
    for(int index=0; index<_components.Length; index++)
    {
        SaveComponent(_components[index]);
    }
}

public void SaveComponent(Component _component)
{
    gameobjectIdentifier = _component.name;
    gameobjectUniqueIdentifier = _component.name;
    PropertyInfo[] _propertyInfo = _component.GetType().GetProperties(System.Reflection.BindingFlags.Instance | System.Reflection.
      BindingFlags.Public);
    FieldInfo[] _fieldInfo = _component.GetType().GetFields(System.Reflection.BindingFlags.Instance | System.Reflection.
      BindingFlags.Public);
    systemDataComponentInstance.Add(new SystemDataComponentInstance(_component.GetType().Name));
    for (int i = 0; i < _propertyInfo.Length; i++)
        systemDataComponentInstance[systemDataComponentInstance.Count - 1].CreateDataInstance(_propertyInfo[i].Name, _propertyInfo[
      i].GetType().ToString(), SystemDataInstance.DataInstanceType.Property);

    for (int i = 0; i < _fieldInfo.Length; i++)
        systemDataComponentInstance[systemDataComponentInstance.Count - 1].CreateDataInstance(_fieldInfo[i].Name, _fieldInfo[i].
      GetType().ToString(), SystemDataInstance.DataInstanceType.Field);
}
```

Listing 7.5: An extract of code used to create the save hierarchy for storing data within the database (see Figure 5.3.4).

discovered during the radio dialogue generation (see Figure 7.4), which creates text responses on the radio depending on if the trainee is looking for help or waiting to supply an answer to the faults identified. When reconstructed normally, the reconstruction is able to reproduce these events fine, as the logic remains the same. However, if an observer skips ahead before instantiate call was made, the object would not be created since the logic is not repeated with the same conditions.

Figure 7.4: The radio dialogue options from the training simulation.

To address this, the solution designed, and partially implemented, includes an extension of Unity's Instantiating system, which introduces an additional layer that overrides existing instantiating calls to pass through this system within the framework. The approach proposed creates a record of the objects being created, whether it is an existing prefab being created from stored assets or a duplication of an existing object already present within the current simulation scene. The recording process relies on the existing Unity engine functionality to handle the recording and reconstruction of variable data. Had this solution been fully implemented, the framework would have reconstructed the instantiating of objects following the same principles as the existing reconstruction methodology. The recorded time instance would be used to determine the creation/deletion of the object along with its individual variable value status, which means that when an observation is skipped ahead or behind when a new gameobject was instantiated, the created object could be configured to exact variables from the recording and deleted as necessary.

7.4 Processed Recording

Processed recording builds upon the data captured during lightweight recording and captures all remaining data not captured during lightweight recording to provide a complete recording of a simulation experience (see Section 5.5). Processed Recording pairs with the reconstruction functionality, allowing the lightweight recording data to drive the reconstruction of user input and non-deterministic data, using reconstruction of data to linearly reproduce simulation modifications. As such, processed recording captures all remaining deterministic variable data.

Processed recording starts by lowering the quality of the simulation to the lowest setting, since the process is expected to be conducted offline, the graphical quality of the simulation is not important. Next, the simulation is searched to determine what data already exists in the database (see Section 6.2) from lightweight recording (see Section 5.4). If an object has already been recorded it is marked to ignore recording requests during processed recording. Afterwords, processed recording re-searches the simulation to identify all remaining variable data that has not been recorded and configures them to be monitored and recorded (during processed recording).

Once all of the data has been configured for recording, a final check is conducted to determine the suitability of data for recording, which is modified to suit the needs of the simulation. These checks determine if any object marked for recording is empty, such as, if the object contains any components or elements which are suitable for recording. Static objects, or objects without any components or variables suitable for recording are ignored. The function for preparing processed recording and do the suitability checks is displayed in Listing 7.6.

However, as processed recording is both a recording and reconstruction component, it is important that while recording outstanding non-deterministic data, the component is also able to consistently reconstruct the user input and non-deterministic data captured from lightweight recording.

7.4.1 Reconstructing Data

The reconstruction architecture is used by both the processed recording and observation components of the framework, with only minor differences separating them. For processed recording, the data is reconstructed using a linear reconstruction methodology that uses the data captured from lightweight recording to to reconstruct the trainee performance experience from the very first frame of the reconstruction. Whereas, the observation component uses the data from both the lightweight and processed recording components to reconstruct the trainee performance using a non-linear reconstruction methodology.

As previously discussed in Sections 7.3 and 5.3.1, the recording components handle and save variable data as type *object* and type string respectively. Casting to these predefined variable types can be performed automatically. However, reconstructing the data is more complicated, since data being reconstructed from the database must be casted back to the original type required by the system. It is not possible to do this automatically, as the casting system requires the variable type to be known and defined. As such, a library system was designed that enables the casting system to explicitly cast data to their original type expected by the game engine (see Section 5.3.3).

```
if (virtualObservationPreprocess == VirtualObservationPreprocess.Yes && virtualObservationPreprocessingMethod == PreprocessingMethod.ALL && virtualObservationStatus != VirtualObservationStatus.Recording)
{
    QualitySettings.SetQualityLevel(0, true); //Lowering Graphical Quality of Experience
    VirtualObservationItemTracker[] _VOITs = FindObjectsOfType<VirtualObservationItemTracker>();
    for (int i = 0; i < _VOITs.Length; i++)
    {
        _VOITs[i]._ignorePreProcess = true;
    }

    List<Transform> _transforms = VirtualObservationTracker.FindObjectsOfTypeAll<Transform>();
    for (int i = 0; i < _transforms.Count; i++)
    {
        bool _found = false;
        for (int x = 0; x < getAllTrackingObjects().Count; x++)
        {
            if (_transforms[i].name == getAllTrackingObjects()[x].name)
                _found = true;
        }

        int objectNameCount = 0;
        for (int x = 0; x < _transforms.Count; x++)
        {
            if (_transforms[x].name == _transforms[i].name)
                objectNameCount++;

            if (objectNameCount >= 2)
            {
                Debug.LogWarning("TWO OR MORE NAMES FOUND: " + _transforms[x].name);
                break;
            }
        }

        if (EligableToSave(_transforms[i].GetComponents<Component>()) && objectNameCount == 1 && _found == false && _transforms[i].gameObject.GetComponent<VirtualObservationItemTracker>() == null && (_transforms[i].gameObject.GetComponentInParent<VirtualObservationTracker>() == null || _transforms[i].gameObject.GetComponent<InteractableObjectInformation>() == null))
        {
            _transforms[i].gameObject.AddComponent<VirtualObservationItemTracker>();
            getVOIH().AddVirtualItemTracker(_transforms[i].gameObject.GetComponent<VirtualObservationItemTracker>().ConfigureForPreprocess());
        }
    }
}
```

Listing 7.6: The launch operation conducted at the start of the processed recording to determine which gameobjects have been previously recorded to filter them out and prepare them for reconstructing using the lightweight recorded data.

7.4.2 Library System

The library system designed and discussed in Section 5.3.3 enables the detection and conversion of data based on the requested casting type. First, a conversion list is created to hold all variable types as their respective string representation. These representations are held within a list, assigning each index of the list to a specific variable type. This conversion list contains all of variable types supported by the library which enable dynamic reconstruction of data. Any variable being reconstructed must be added to this list. By default, the library system is designed to automatically support predefined variable types, however, additional variables can be added to the list, allowing easy expansion of data supported by the library system. Listing 7.7 shows the incorporation of predefined variable types, along with a developer defined types demonstrating that the library system supports variables of any type.

While the library system resolves most of the issues for reconstructing data, it still requires some effort on the part of the developer to add and expand to the list as necessary for the individual needs of the simulation. However, with the library system, the developer input to extend the supported variable types is isolated to a single area within the development environment of the framework, with only three lines of code necessary to expand the variables supported by the library system.

To expand upon the list, the variable type name that represents a variable type is added to the list of '*ConversionTypes*', this assigns the variable type as an index value. Next, within the conversion function, an additional *IF* statement is included to compare the saved variable type name against the checked index value of the variable passed into the function. To convert the variable back to its original type, a final piece of code is included which returns the value of the converted variable using the *Convert.Change* type function. Because the index of the conversion type is known, the line can be written to support the new variable type. In return, the new variable type is now supported for dynamic reconstruction by the framework.

7.4.3 Reproduction of Simulation Data

With the library system providing the means for data to be explicitly cast to their required type for the game engine, the framework is capable of dynamically reconstructing data.

The reconstruction of data is split into two sectors of the Unity engine executing order (see Section 7.2), replicating the recording process from Section 7.3.2 for Fixed Update and Late Update, once again focusing on the differences for user input and object recording data.

A key part of the reconstruction process is shown in Listing 7.8, shows the pseudocode for the framework logic for reconstructing the captured data in a forward or rewind direction, depending on the current mode of reconstruction. The forward reconstruction generates the actions in a logical order based on their time of recording to reproduce the exact actions of the trainee from start to finish, whereas, the rewind directions reproduces the trainee actions in a reverse, playing the actions backwards until the start of the simulation is reached.

```
//List of library defined variable types used to convert saved string data back to original variable type.
public static string[] ConversionTypes = {
    "String", // 0
    "Float", // 1
    "Int32", //2
    "Bool", // 4
    "Double", //5
    "UInt32", // 9
    "VRInputManager+ButtonStatus" // 11,
};

//Conversion process can be called by suppling the string variable data and the string original variable type.
public static object ConvertData(string variableData, string variableType)
{
    if (Includes(variableType, ConversionTypes[4]))
        return Convert.ChangeType(variableData, typeof(bool));
    if (Includes(variableType, ConversionTypes[0]))
        return Convert.ChangeType(variableData, typeof(string));
    if (Includes(variableType, ConversionTypes[1]))
        return Convert.ChangeType(variableData, typeof(float));
    if (Includes(variableType, ConversionTypes[9]))
        return Convert.ChangeType(variableData, typeof(uint));
    if (Includes(variableType, ConversionTypes[2]))
        return Convert.ChangeType(variableData, typeof(int));
    if (Includes(variableType, ConversionTypes[5]))
        return Convert.ChangeType(variableData, typeof(double));
    if (Includes(variableType, ConversionTypes[11]))
        return (VRInputManager.ButtonStatus)System.Enum.Parse(typeof(VRInputManager.ButtonStatus), variableData);

    //throw error - alert developer which variable conversion (value and type) is missing!
    return null;
}
```

Listing 7.7: The library system which explicitly casts data. This pseudocode shows the process for the libary system casting the string data saved back to its original format using the predefined ConversionTypes.

```
public void Rebuild(float time, VirtualObservationRebuilder.ReconstructionReplayMode reconstructionReplayMode, bool forceSimulate =
        false)
{
    if (reconstructionReplayMode == VirtualObservationRebuilder.ReconstructionReplayMode.Forward)
    {
        systemDataClass._enabled.ForwardRebuild(ref _gameObjectRef, time, _ignoreRebuildIndex);
        this.transform.position = systemDataClass._position.ForwardRebuild(_gameObjectRef.transform.position, time,
      _ignoreRebuildIndex);
        systemDataClass._rotation.ForwardRebuild(ref _gameObjectRef, time, _ignoreRebuildIndex);
    }
    else if(reconstructionReplayMode == VirtualObservationRebuilder.ReconstructionReplayMode.Rewind)
    {
        systemDataClass._enabled.RewindRebuild(ref _gameObjectRef, time, _ignoreRebuildIndex);
        this.transform.position = systemDataClass._position.RewindRebuild(_gameObjectRef.transform.position, time,
      _ignoreRebuildIndex);
        systemDataClass._rotation.RewindRebuild(ref _gameObjectRef, time, _ignoreRebuildIndex);
    }

    _indexCount = systemDataClass._rotation.Index;
    positionIndex = systemDataClass._position.Index;
    _currentTimer = time;
}
```

Listing 7.8: Pseudocode of transform data being reconstructed.

There are two stages of the reconstruction process. The first stage reproduces the transform information of the recorded data, with the second stage reproducing the logic of the variable values. Transform data includes all data that determines the position, rotation, scale and active state of an object. As such, during Fixed Update all recorded transform data for objects is reconstructed. The remaining data is considered logic data, since it derives from components and additional variables from an object.

Transform data is reproduced first in the Fixed Update cycle before logical data is reproduced as it provides an opportunity for the engine to update physics calculations which drive many of the interaction systems in virtual reality simulations. For example, to reconstruct the physics attributes of a user picking up an object, the object must match the exact transformation instance that allows the physics system embedded within the game engine to register the object. In this instance, the simulation will detect a collision, identifying that the object is within range for pickup and only requires the user input to initiate the action. As a consequence, logical data is produced later in the Late Update stage as it requires the input to replicate the simulation logic. If this logic was not applied and the physics calculations had not been completed, input would be reproduced too early and the detection of objects from collision may not be detected, causing divergence in the reconstruction. As such, the transform data for both the trainee and the object (technically all objects in the simulation environment) should be reconstructed first, followed by unity's internal physic calculation, with input and logic data being reconstructed last.

With data reproduced in this manner, the reconstruction process is iterative, with all recorded data monitored every-frame to determine if the data is suitable for reconstruction. The framework loops the recording data during each reconstruction stage to determine if the next reconstruction time matches the current time of observation, if so the data is reconstructed, using the library system to reconstruct the data through casting, converting string variables back to their original type (see Section 7.4.2). When a variable reconstruction index is complete all data recorded has been reconstructed, the variable data is marked as complete so it is ignored on subsequent reconstruction calls. The complete marking is removed if the reconstruction direction is altered, since the variable data is then reconstructed in a different direction (forward or reverse). Listing 7.9 shows an extract of code used to reconstruct an individual data instance, which includes use of try and catch conditions to provide redundancy in-case of any unexpected failures during the operation.

On top of the normal reconstruction process, the framework incorporates a skipping functionality, in which the data is reproduced at a faster rate than is possible during normal a simulation. This feature was implemented as part of the replay feature (see Section 6.3.3), which was designed to meet the requirements R1, R2 and R4 (see Section 4.2) from the requirement collection study with assessors in Section 4.1. This functionality is used for the Observation and Assessment components and enables the reconstruction to reproduce data at a faster rate than was experienced in the recorded simulation The reconstruction uses all data from the database to ensure that the reconstruction of the simulation is kept in sync, enabling a non-linear reconstruction to be achieved.

```

public void ChangeDataInstance(object component, float time, string _parentVariableName = "null", object parentObject = null)
{
    if (variableValue.Count == 0) return;

    if (dataInstanceType == SystemDataInstance.DataInstanceType.Field)
    {
        object varData = null;
        if (AdvancedTrackingDataObjects.ValidData(variableType) || AdvancedTrackingDataObjects.ValidData(component.GetType().
  GetField(variableName).FieldType.ToString()))
        {
            try
            {
                varData = VirtualObservationAdvancedData.ConvertData(variableValue[Index], variableType);
            }
            catch { Debug.Log(variableName + " " + variableType + " " + variableValue.Count + " " + index); }

            component.GetType().GetField(variableName).SetValue(component, varData);
        }
    }
}
```

Listing 7.9: Extract of pseudocode used to reconstruct the data values within the Unity engine (see Section 5.3).

```
IEnumerator FastProcessReconstructionRewind(float newTime)
{
    lockControls = true;
    isSkimmingSimulation = true;
    VirtualObservationRebuilder.ReconstructionReplayMode tempReplayMode = reconstructionReplayMode;
    reconstructionReplayMode = VirtualObservationRebuilder.ReconstructionReplayMode.Rewind;
    yield return new WaitForSeconds(0.05f);

    VOSDataStream.ForceTimeReconstructionRewind(global_timer, newTime);
    VOSDataStream.ForceTimeReconstructionRewind(global_timer, newTime, false);

    yield return new WaitForSeconds(0.05f);
    getVOIH().SetReconstructionTime(global_timer, newTime, VirtualObservationRebuilder.ReconstructionReplayMode.Rewind, true);
    yield return new WaitForSeconds(0.05f);
    getVOIH().SetReconstructionLateUpdate(newTime, global_timer, VirtualObservationRebuilder.ReconstructionReplayMode.Rewind);
    global_timer = newTime;

    yield return new WaitForSeconds(0.1f);
    GetControllerHeldObjects();
    yield return new WaitForSeconds(0.5f);
    lockControls = false;
    Pause();
    skippingValueFrames = 0;
    isSkimmingSimulation = false;
    yield break;
}
```

Listing 7.10: Pseudocode used to fast process the simulation to skip to a predefined time in the reconstructed trainee performance. This functionality provides non-linear reconstruction, allowing the simulation data captured from the recording components to be reconstructed at any time period from within the trainee performance.

Listing 7.10 shows the code used to provide non-linear reconstruction of simulations. This code modifies the reconstruction time of the simulation by reconstructing and syncing all data to match the new time requested by the assessor operating the framework. This operates under the IEnumerator classification to prevent the operation from freezing the interface or other features of the game engine while the data is reconstructed.

With reconstruction able to reproduce trainee performances, it is important that the framework provides the technical capabilities for an assessor to observe, assess and provide feedback to a trainee through direct observation.

7.5 Observation, Authentic Assessment and Contextual Feedback Functionality

With the technical architecture of the framework implementing all necessary functionality to capture and reproduce trainee performances, the remaining sections of this chapter focuses on the most technically challenging aspects of the components, tools, and features that aim to make use of the reconstruction so that assessors can observe, assess and provide feedback for room-scale trainee VR performances.

As such, the observation component focuses on the implementation of the observation of the free camera tool (see Section 6.3.2) and user interface. As the recording and reconstruction stages provide the necessary functionality to record and reconstruct simulations, the observation stage allows an assessor to directly observe the trainee to make sense of the data. This is done by enabling assessors to directly observe and monitor the trainee actions during the reproduced training simulation, but first it is important create a user interface that enables assessors to engage with the reconstructed simulation.

User Interface

The user interface was designed to scale to the user resolution and screen size, allowing all elements held within the UI to scale to suit the screen resolution (Figure 7.5). Because the framework implementation is universal to all platforms supported by the Unity engine, the interface was designed to be universally compatible for use on mobile and web platforms as was demonstrated in (Howie and Gilardi, 2020). The interface could support any deployment platform by adapting and scaling to suit the resolution of display. However, for the purpose of the evaluation in Chapter 8 only the desktop version of the framework was used.

Inspired by video editing interfaces, the augmented timeline (see Section 6.5.2) and controls are positioned to resemble and operate similar to video editing suites (Magix, 2021). Likewise, the incorporation of additional assessment tools and features was partly inspired by the work of Raij and Lok (2008) and Stone, Snell, and Cooke (2016), which demonstrated overlaid interfaces within a simulation environment to display additional information that would otherwise be difficult to view. Figure 7.6, presents the implementation of the summative performance indicators interface designed in Section 6.5.1.

7.5.1 Free Camera Tool

To meet the identified the requirements of R1, R2, R5 and R6 from assessors (see Section 4.2), the free camera tool designed in Section 6.3.2 was implemented into the framework. The camera controls were designed to replicate existing methodologies (Stone, Snell, and Cooke, 2016; Raij and

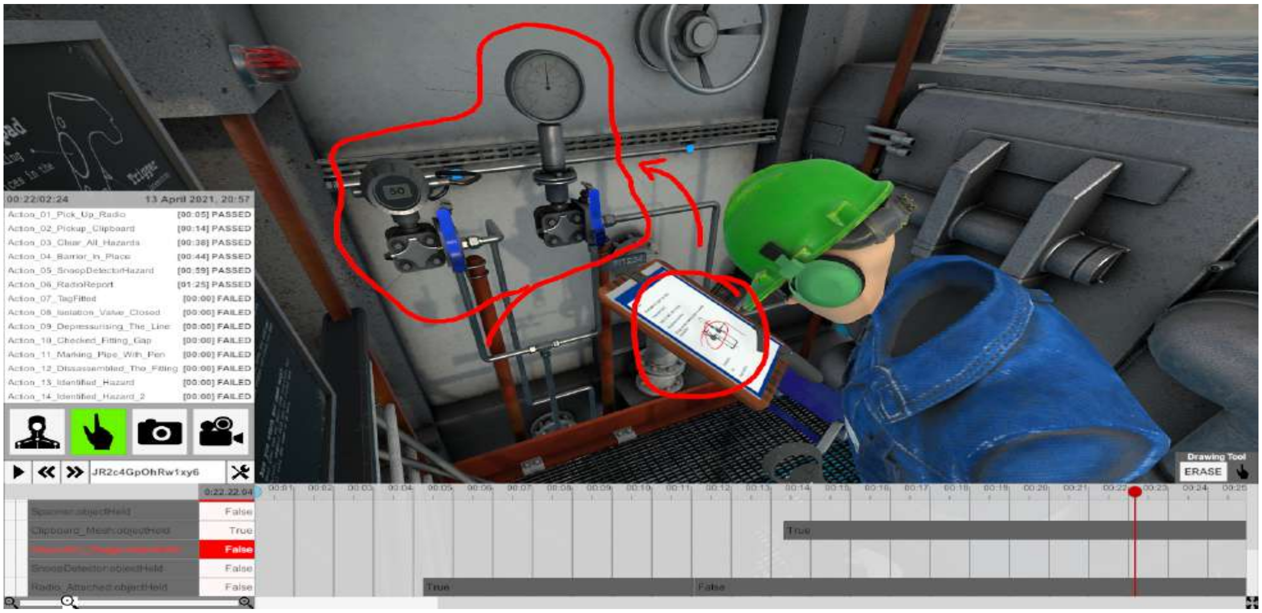


Figure 7.5: The interface of the framework, showing the augmented timeline and drawing tool features. See sections 6.5.2 and 6.6.3 for design.

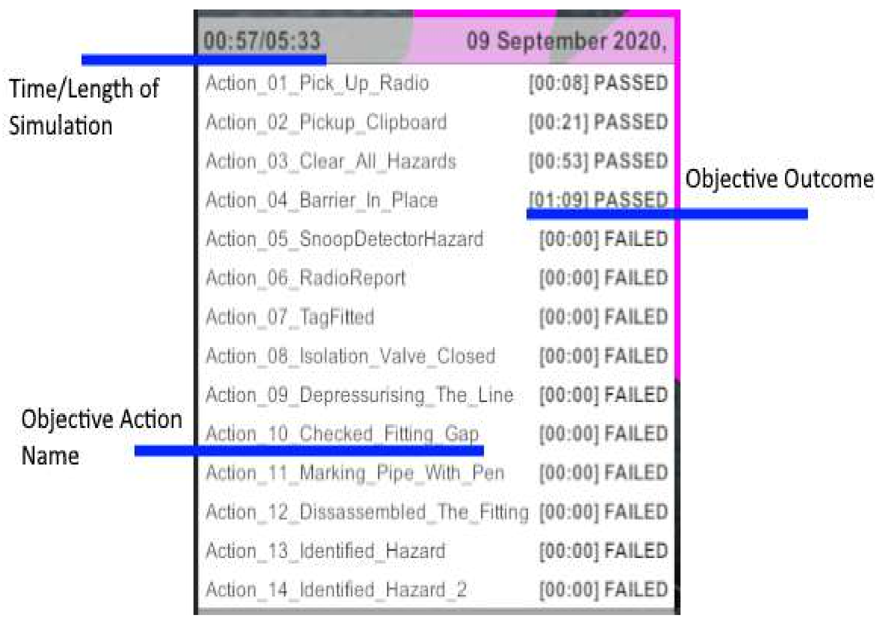


Figure 7.6: Implementation of the summative performance indicators tool. See section 6.5.1 for design.

Lok, 2008) for free camera tool designed in Section 6.3.2 (Figure 7.7). By designing the controls based on existing implementations, including Unity's default camera controller, it was hoped that the controls would be familiar to assessors who were operating the simulation. However, two challenges were faced during development of the camera controls. Firstly, pausing the reconstruction of the simulation caused the camera functionality to stop, and secondary, remote-screen sharing applications prevented the camera from recognising the remote assessors mouse movement.

Figure 7.7: Presentation of the free camera tool represented within the simulation environment. This figure shows the position of the camera within the 3D world-space of the reconstructed simulation and a preview of the camera perspective in the bottom right. This previous shows the perspective of observation an assessor would be viewing the trainee from.

The first challenge was overcome by keeping the *'Time.scale'* of the engine above zero, so that controls were always available for the observer to use. *'Time.scale'* is the value of time within the unity engine that is modified each second, when at zero, time stops and the simulation pauses, which in turn caused user control of the camera to stop. The framework incorporated its own time keeping system into the reconstruction process to separate it from Unity's default method to overcome this issue.

The second issue only revealed itself during testing for remote evaluation of the framework using remote-screen sharing applications (see Section 3.8.2). Due to the implementation for how remote screen-sharing applications provide their control functionalities, the Unity game engine refused to accept the input from the default position of the mouse using its standard API implementation. The controls to operate the camera which enabled the mouse control (rotation) was modified to take accept the raw values of input rather than using the in-build API within the Unity Engine. The raw values were taken from the screen-coordinates of the mouse pointer and modified each frame to determine the differences in movement. This process manually updated the mouse co-ordinates, which enabled the camera to function. With the assessor now able to manually control the camera and monitor the user interface, the final step was to populate the interface with the augmented timeline generation of variable data.

```
    if (_found)
    {
        GameObject _componentPrefab = null;
        if (vouiDataPresentationMode != VOUIDataPresentationMode.Basic)
        {
            _componentPrefab = CreateGameComponentName(_gameObjectPrefab.GetComponent<VirtualObservationUIExpansion>(), VOT.
    getVOSDataStream().SystemDataInstance[i].systemDataComponentInstance[x]);
        }
        for (int z = 0; z < VOT.getVOSDataStream().SystemDataInstance[i].systemDataComponentInstance[x].systemDataInstance.Count; z
    ++)
        {
            if (vouiDataPresentationMode == VOUIDataPresentationMode.Basic)
            {
                CreateInstanceForList(_heldGameObjectCurrentValueData, _heldGameObjectTimelineData, VOT.getVOSDataStream().
    SystemDataInstance[i].systemDataComponentInstance[x].systemDataInstance[z], VOT.getVOSDataStream().SystemDataInstance[i].
    gameobjectIdentifier, VOT.getVOSDataStream().SystemDataInstance[i], VOT.getVOSDataStream().SystemDataInstance[i].
    systemDataComponentInstance[x]);
                CreateItemForList(_heldGameObjectCurrentValueData, _heldGameObjectTimelineData, VOT.getVOSDataStream().
    SystemDataInstance[i].systemDataComponentInstance[x].systemDataInstance[z], VOT.getVOSDataStream().SystemDataInstance[i].
    gameobjectIdentifier, VOT.getVOSDataStream().SystemDataInstance[i], VOT.getVOSDataStream().SystemDataInstance[i].
    systemDataComponentInstance[x]);
            }
            else if(vouiDataPresentationMode == VOUIDataPresentationMode.Advanced || vouiDataPresentationMode ==
    VOUIDataPresentationMode.Developer)
            {
                CreateInstanceForList(_gameObjectPrefab.GetComponent<VirtualObservationUIExpansion>(), _componentPrefab.
    GetComponent<VirtualObservationUIExpansion>(), VOT.getVOSDataStream().SystemDataInstance[i].systemDataComponentInstance[x].
    systemDataInstance[z], VOT.getVOSDataStream().SystemDataInstance[i], VOT.getVOSDataStream().SystemDataInstance[i].
    systemDataComponentInstance[x]);
                CreateItemForList(_gameObjectPrefab.GetComponent<VirtualObservationUIExpansion>(), _componentPrefab.GetComponent<
    VirtualObservationUIExpansion>(), VOT.getVOSDataStream().SystemDataInstance[i].systemDataComponentInstance[x].
    systemDataInstance[z], VOT.getVOSDataStream().SystemDataInstance[i], VOT.getVOSDataStream().SystemDataInstance[i].
    systemDataComponentInstance[x]);
            }
        }
    }
```

Listing 7.11: An extract of pseudocode used to check objects or data suitability for being displayed on the augmented timeline interface.

7.5.2 Augmented Timeline

The augmented timeline (see Section 6.5.2) is the final technical implementation of the framework that is considered technically challenging, requiring the algorithms and code to be discussed and documented. The augmented timeline generation utilised the data captured from both recording components (see Section 5.3) and stored in the database (see Section 6.2) to generate a timeline for all variable data within a simulation (see Section 6.5.2). This was created to meet requirement R4 (see Section 4.2) from the study conducted with assessors in Section 4.1.

To generate the timeline, the framework loads the trainee data-set for reconstruction (after processed recording had been completed), once all data was loaded, the augmented timeline architecture system scans through all data looking for predesignated variable names that were marked for display on the timeline. It was envisioned that the variable data would be dynamically added to the timeline, however due to time constraints, the implementation is only able to use predesignated variables during initial creation of the timeline. This means that variables that are to be displayed on the augmented timeline have to be set by a developer before the reconstruction operation takes place. When a predesignated variable was discovered during the initial launch of the simulation during reconstruction, the variable was added to a list of stored variables that was assigned for display on the timeline. An extract of code used to check suitability of data for display on the augmented timeline is shown in Listing 7.11.

When the data is being configured for reproduction for direct observation or assessment by an assessor, the augmented timeline is generated at run-time and is embedded onto the user interface from Section 7.5. Each variable that is stored in the list for display is provided with an individual row on the timeline, with the contents of the row filled with information blocks of data that are scaled to reflect their relative time of modification. For example, a variable value that was true for 2 seconds, would have the information block scaled to fit 2 seconds in the augmented timeline. Figures 7.8 and 7.9 show the implementation of the timeline described in Section 6.5.2.

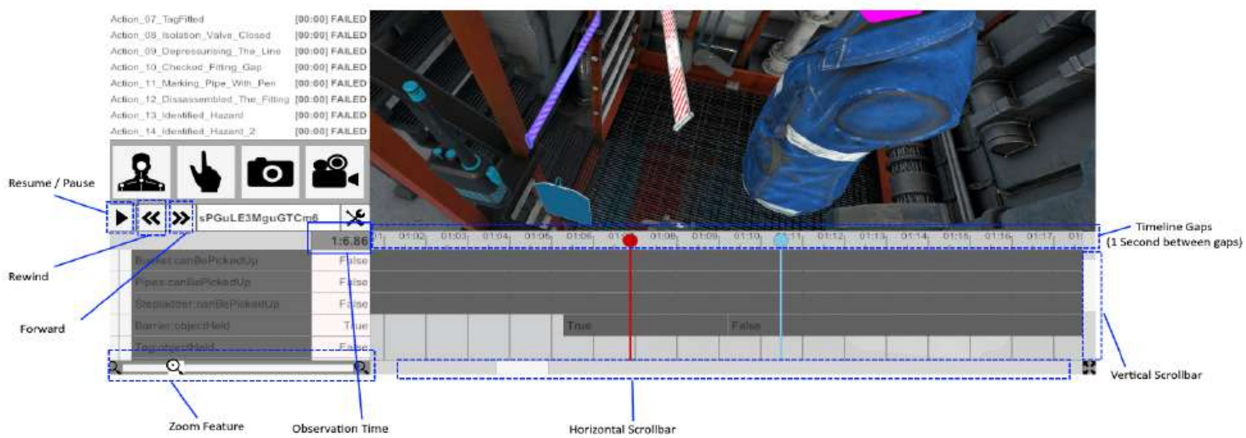


Figure 7.8: The tools, control and augmented timeline with labels that describe the implementation. See section 6.5.2 for the design of the augmented timeline.

Listing 7.12 provides code of the augmented timeline implementation for displaying the predesignated variables. First, the framework determines if the variable data is part of the predesignated variable list, if not the variable is skipped. Otherwise a new row is created and made active and visible. All information relevant to the variable being displayed on the row is assigned, including the name and initial value. If the variable contains no history, the variable name is marked as red to indicate no variable modifications were detected. For variables that do have modification history associated with them, each individual variable modification is assigned an individual block on the variables row in the timeline, with the modification value of the variable embedded into the block. Each individual value on the block is scaled to represent the length of the variable modification, which is synced against the length of the recording (see Figures 7.8 and 7.9). By following this implementation, the technical implementation for the framework augmented timeline can be easily replicated.

```
public void CreateItemForList(VirtualObservationUIExpansion _gameObjectClass, VirtualObservationUIExpansion _componentClass,
    SystemDataInstance _systemDataInstance, SystemDataClass _systemDataClass, SystemDataComponentInstance
    _systemDataComponentInstance)
{
    if (!_matchExactVariablesOnTimeline)
    {
        if (!ValidDataToDisplay(_systemDataInstance.variableName))// || _systemDataInstance.variableValue.Count == 1)
            return;
    } else {
        if (!ValidDataToDisplay(_systemDataInstance, _systemDataClass, _systemDataComponentInstance))
            return;
    }

    GameObject rowItem = CreateItemForRow(ref _componentClass._heldGameObjectNames, _systemDataInstance.variableName);
    _timeLineVerticalList.Add(Instantiate(_timerVerticalListItem, _timerVerticalListItem.transform.position, _timerVerticalListItem
      .transform.rotation, _timerVerticalListItem.transform.parent));
    rowItem.GetComponent<VirtualObservationUIExpansion>().SetVOTimelineGroup(_timeLineVerticalList[_timeLineVerticalList.Count -
      1].GetComponent<VirtualObservationUITimelineGroup>());
    _componentClass._heldGameObjectTimelineData.Add(_timeLineVerticalList[_timeLineVerticalList.Count - 1]);
    _componentClass._heldGameObjectTimelineData[_componentClass._heldGameObjectTimelineData.Count - 1].SetActive(false);
    GameObject _varDataValue = CreateCurrentDataInstanceValue(ref _componentClass._heldGameObjectCurrentValueData,
      _systemDataInstance.variableName);
    _varDataValue.GetComponentInChildren<Text>().resizeTextForBestFit = true;
    _varDataValue.GetComponentInChildren<Text>().resizeTextMaxSize = _varDataValue.GetComponentInChildren<Text>().fontSize;
    _timeLineContent.Add(new List<GameObject>());
    _timeLineVerticalList[_timeLineVerticalList.Count - 1].name = _systemDataInstance.variableName;
    int prefabIndex = 0;

    for(int i=0; i<_systemDataInstance.variableValue.Count; i++)
    {
        _timeLineContent[_timeLineContent.Count - 1].Add(Instantiate(TimelineContentPrefab, TimelineContentPrefab.transform.
      position, TimelineContentPrefab.transform.rotation, _timeLineVerticalList[_timeLineVerticalList.Count - 1].transform));
        GameObject _createdInstance = _timeLineContent[_timeLineContent.Count - 1][_timeLineContent[_timeLineContent.Count - 1].
      Count - 1];
        if(i==0)
            _createdInstance.GetComponent<VirtualObservationUIInstance>().Create(getTimelineDuration(_systemDataInstance, i),
      _systemDataInstance.variableValue[prefabIndex], this, vouiDataPresentationMode, true);
        else
            _createdInstance.GetComponent<VirtualObservationUIInstance>().Create(getTimelineDuration(_systemDataInstance, i),
      _systemDataInstance.variableValue[prefabIndex], this, vouiDataPresentationMode, false);

        _UITimelineInstances.Add(_createdInstance.GetComponent<VirtualObservationUIInstance>());
        _createdInstance.SetActive(true);

        _timeLineVerticalList[_timeLineVerticalList.Count - 1].GetComponent<VirtualObservationUITimelineGroup>().Create(
      _varDataValue.GetComponentInChildren<Text>(), _systemDataInstance.variableValue[i], getMinTimerValue(_systemDataInstance, i),
       getMaxTimerValue(_systemDataInstance, i));
        _timeLineVerticalList[_timeLineVerticalList.Count - 1].GetComponent<VirtualObservationUITimelineGroup>()._textReference.
      text = _systemDataInstance.variableName;
        _createdInstance.GetComponent<VirtualObservationContentInformation>().Create(this, _systemDataInstance.variableName,
      _systemDataInstance.variableValue[i]);
        prefabIndex++;
    }
    _timeLineVerticalList[_timeLineVerticalList.Count - 1].SetActive(true);
}
```

Listing 7.12: Pseudocode used to generate augmented timeline instances for objects displayed on the augmented timeline.

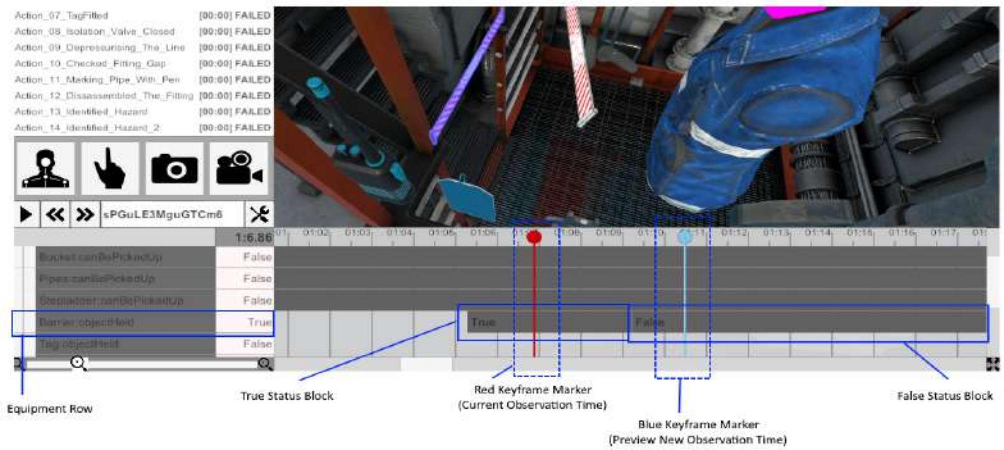


Figure 7.9: The augmented timeline with labels that describe the implementation for blocks of true or false data, which represent the key-points of the simulation, meeting requirement R4 from Section 4.2. See section 6.5.2 for the design of the augmented timeline.

7.6 Chapter Summary

This chapter presented the technical implementation of the framework that was designed in Chapters 5 and 6 to meet the functional requirements of assessors (see Section 4.2) identified from the requirement collection study in Section 4.1.

The framework implemented in this chapter allows an evaluation to be conducted with assessors to answer RQ6 and RQ7. In the following Chapter 8, the evaluation determined if the framework and its components, tools, and features meet the requirements of assessors (see Section 4.2) and identify how the framework compares against existing methods for conducting authentic assessment and contextual feedback.

Chapter 8

Evaluation of The Framework

This chapter answers the final research questions of this work RQ6 and RQ7, which ask *"Are the tools and features provided in the framework suitable for conducting authentic assessment and generating contextual feedback of room-scale virtual reality training simulations?"* (RQ6) and *"How useful is the framework compared to state-of-the-art methods for conducting authentic assessment and generating contextual feedback of room-scale virtual reality training simulations?"* (RQ7).

This chapter discusses the evaluation of the framework that was designed and developed in Chapters 5, 6 and 7, based on the requirements identified from the study with industrial training assessors for conducting authentic assessment and generating feedback for room-scale VR training simulations (see Chapter 4). The methodology summarised in Section 8.1. Section 8.2 presented the thematic analysis process used to generate and form the final themes for each factor, presenting extracts and showing the evolution of data, which formed the key findings of the evaluation. These key findings are presented in Section 8.3, formed of the themes: *Audio and Visual Cues*, *Assisted Assessment and Feedback*, *Point of Observation during Assessment* and *Follow-up Dialogue Session*. Section 4.2 evaluates the tools and features of the framework designed in Chapter 6 to determine if they satisfy the requirements of assessors (see Section 4.2). Limitations of the framework are discussed in Section 8.5. Finally, the last Section 8.6, concludes with a summary of the findings and their implications for the final two research questions (RQ6 and RQ7) of this work.

8.1 Methodology

As discussed in Chapter 3, the evaluation was conducted with 13 participants recruited from industry training and academic assessors, 6 participants from industry and 7 from academia. The sample size was considered suitable following Malterud, Siersma, and Guassora's (2016) "information power" theory, which proposed findings from a targeted sample of participants can produce better comprehensive data when accounting for participants extensive knowledge within the subject area when using quality interview dialogues (see Section 3.4). The evaluation was designed as a repeated measure within groups experimental (within-subjects) designs, with participants using 3 factors for the evaluation: the framework, analytical data and video assessment factors (see Section 3.8.1). The order of factors used were assigned to participants using a Latin Square Order (Lazar, Feng, and Hochheiser, 2017) to avoid bias in participation experimentation order (Table 8.1). The evaluation used a qualitative study, which involved questionnaires to guide discussions during semi-structured interviews (see Section 3.8), the resulting data was analysed following the thematic analysis process described by Braun and Clarke (2006) (see Section 3.5.1). A commercial room-scale VR training simulation was used as the evaluation scenario (see Section 3.8). Participants conducted assessment and generated feedback for two VR trainee performance data-sets (A and B) (see Section 3.8). Due to COVID19 restrictions, the evaluation was hosted on the author's PC and streamed to participants over Microsoft Teams (MS) remote-screen sharing feature, giving the remote participants full control over the interactions of the methods using MS teams remote-control tool (see Section 3.8.2).

Table 8.1: Latin Squares Order (Lazar, Feng, and Hochheiser, 2017) used during the study.

Order of Factor use by Participant			
ID	First	Second	Third
1	The Framework	Video Assessment	Analytical Data
2	Analytical Data	The Framework	Video Assessment
3	Video Assessment	Analytical Data	The Framework
4	The Framework	Video Assessment	Analytical Data
5	Analytical Data	The Framework	Video Assessment
6	Video Assessment	Analytical Data	The Framework
7	The Framework	Video Assessment	Analytical Data
8	Analytical Data	The Framework	Video Assessment
9	Video Assessment	Analytical Data	The Framework
10	The Framework	Video Assessment	Analytical Data
11	Analytical Data	The Framework	Video Assessment
12	Video Assessment	Analytical Data	The Framework
13	The Framework	Video Assessment	Analytical Data

8.2 Thematic Analysis Theme Generation

Thematic Analysis was used to analyse the data from the evaluation. Braun and Clarke's (2006) processes was followed to generate the final thematic maps for each of the factors, *the framework* (see Section 8.1), *analytical data* (see Section 8.1) and *video assessment* (see Section 8.5).

The initial coding of the transcribed documents to the final thematic maps presented in this thesis followed Braun and Clarke's (2006), iterative process, with frequent modifications, re-coding, and re-generating of the thematic maps. After the data was transcribed, each data-set was analysed twice to ensure familiarity with the data, followed by coding sections to form the initial themes. The processing of coding the data was conducted systematically throughout the data-set, giving equal attention to each data item to identify interesting connections of data which would eventually form the structure of the initial themes. These initial themes represented repeating patterns within the data-set, forming an overall basis for further analysis. The themes were reviewed and re-generated as the themes was organised to support the findings. During this process, the initial themes were generated and then reviewed against the data, iteratively modifying and merging themes to create a coherent and meaningful theme that represented the findings of the data.

Initially, the themes were generated by combining a set of different coded sections of data to form an overarching initial theme. At a broader level, these themes were generated to represent an overall representation of the codes and their relation to each other. After the initial thematic map was generated, the themes were reviewed against the data to determine how they reflect on the data-set as a whole. Once the themes provided an accurate representation of the data which fit together and presented an overall structure for the data representation, the analysis moved to define the themes. The themes were named to define the structure and meaning of the data, summarising the overall findings from a large data-set of coded information. Only once it was felt that the data presented in the thematic maps represented the findings from the data, the final thematic maps were generated. The final themes were categorised into sections of Usefulness, Usability and Problems. These categories allowed each theme attached to provide a narrative story, which presents the essence of the findings from the data related to each category.

In this chapter, extracts form the collected data are used as evidence to justify the analysis conclusions that provides confidence in the findings. The themes from thematic analysis are presented using a narrative structure to provide the reader with the interesting and relevant findings from the work in a format that is easy to digest.

8.2.1 Framework Factor Thematic Phases

Figure 8.1 shows the thematic process used to determine the final themes of the evaluation. This process grouped the themes onto the factors which reflected the aim of answering RQ6 and RQ7. Figure 8.2 presents the evolution of themes from the data, showing a section of the analysis conducted, including cues, initial themes and final themes. While this was an iterative process, Figure 8.2 attempts to simplify the approach for easier visual clarity.

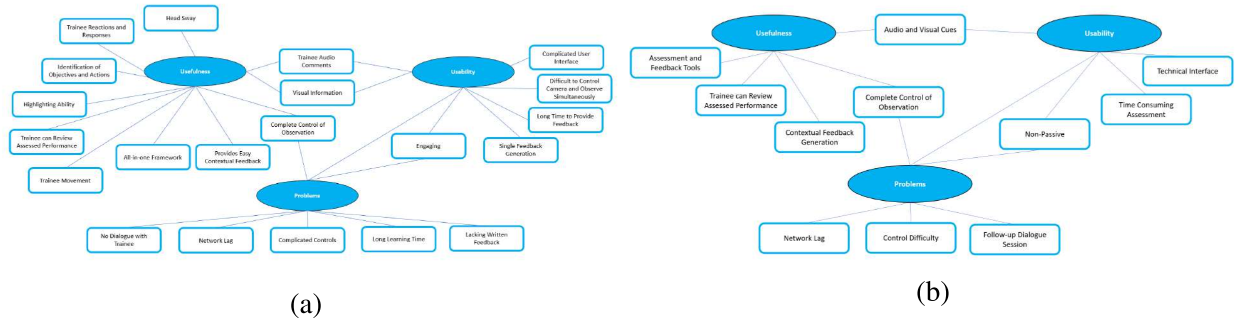


Figure 8.1: (a) Initial Framework thematic map, (b) Final Framework thematic map. Full resolution images of the thematic maps are located in Appendix E.

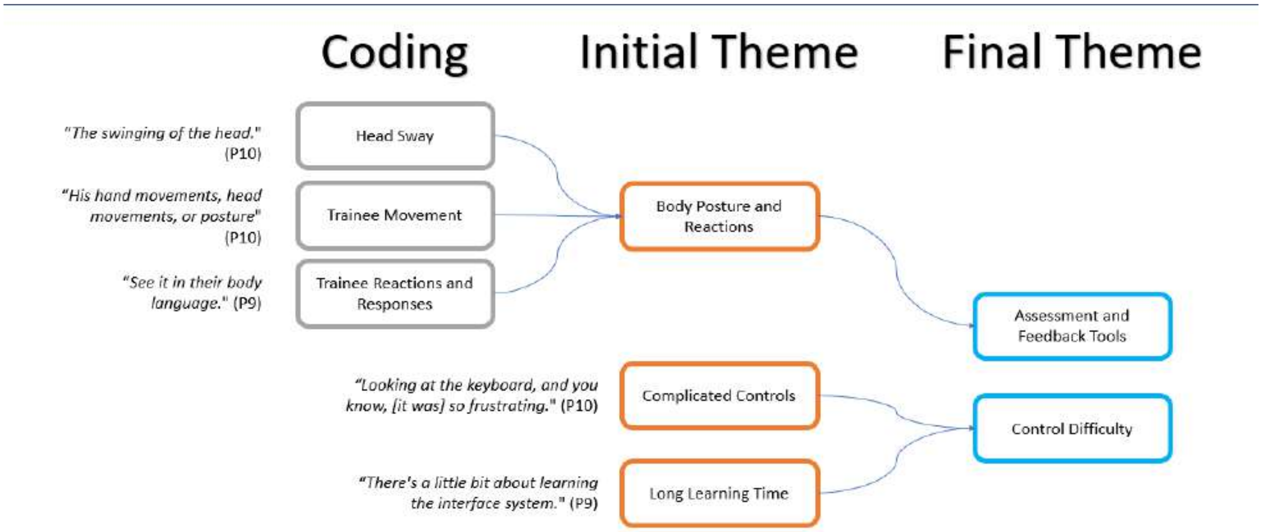
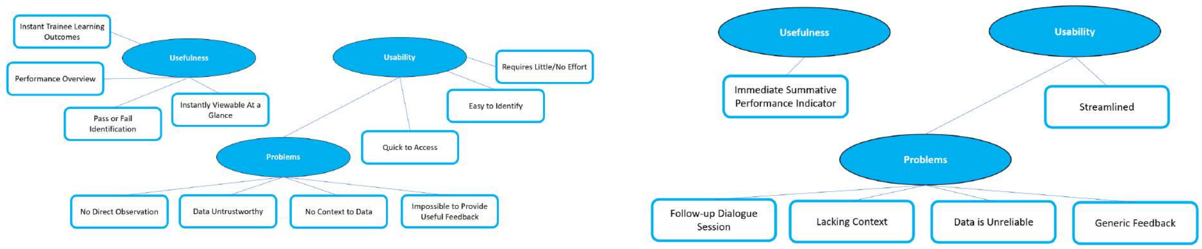


Figure 8.2: An extract of the thematic process conducted for the framework, showing the evolution of themes of *Assessment and Feedback Tools* and *Control Difficulty* from cues, to initial themes and final themes.

8.2.2 Analytical Data and Video Assessment Factors Thematic Phases

The thematic phases of the data was identical for the analytical data and video assessment factors. Figure 8.3 shows the comparison for initial and final themes generated via thematic analysis for analytical data, with Figure 8.5 doing the same for the video assessment factor. Figures 8.4 and 8.6 presents an extract of the analysis conducted to generate the cues, initial themes and final themes for each thematic map, for analytical data and video assessment factors.



(a) (b)

Figure 8.3: (a) Initial Analytical Data thematic map, (b) Final Analytical Data thematic map. Full resolution images of the thematic maps are located in Appendix E.

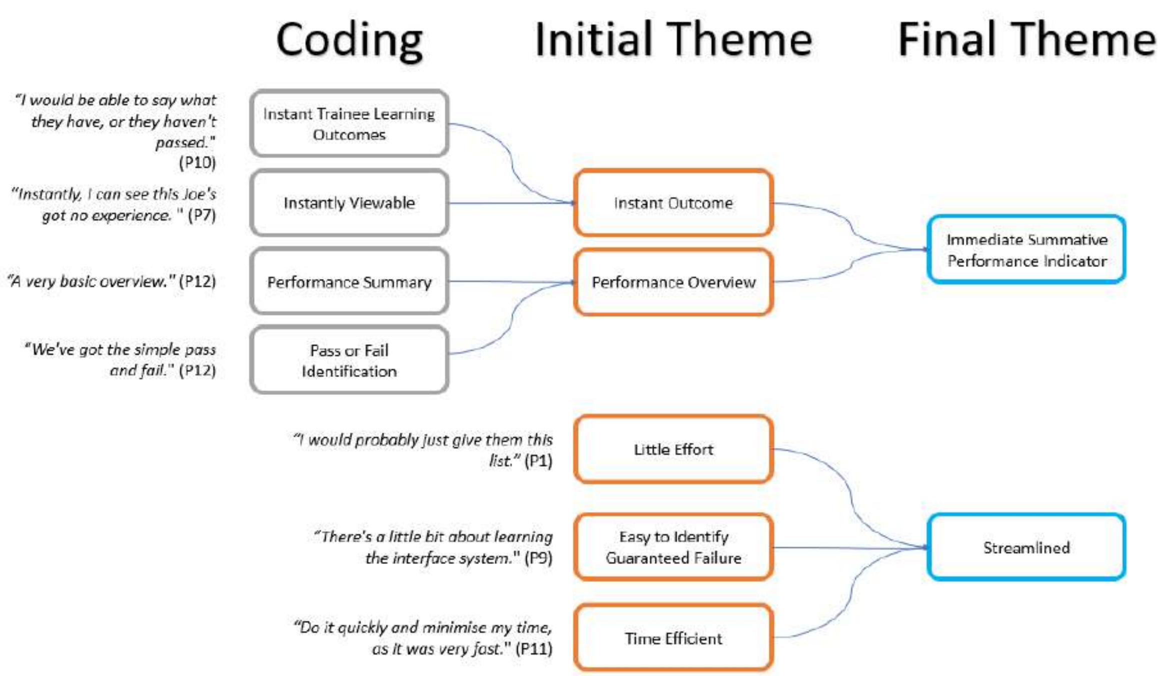


Figure 8.4: An extract of the thematic process conducted for analytical data factor, showing the evolution of themes of *Immediate Summative Performance Indicators* and *Streamlined* from cues, to initial themes and final themes.

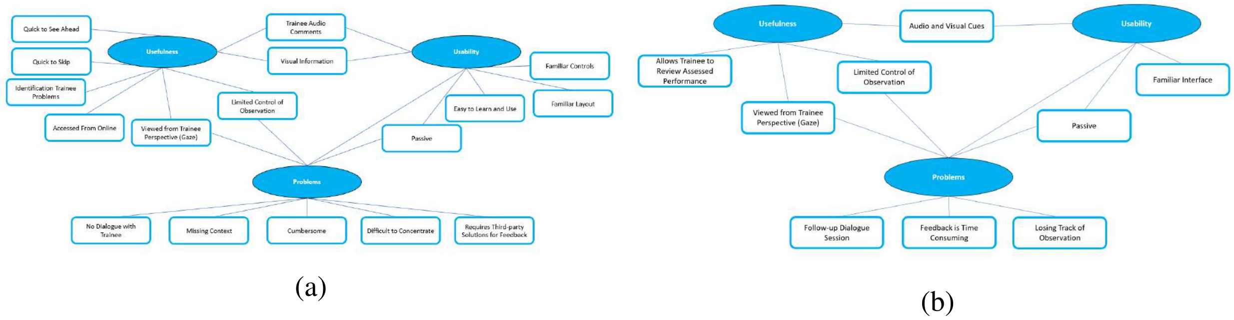


Figure 8.5: (a) Initial Video Assessment thematic map, (b) Final Video Assessment thematic map. Full resolution images of the thematic maps are located in Appendix E.

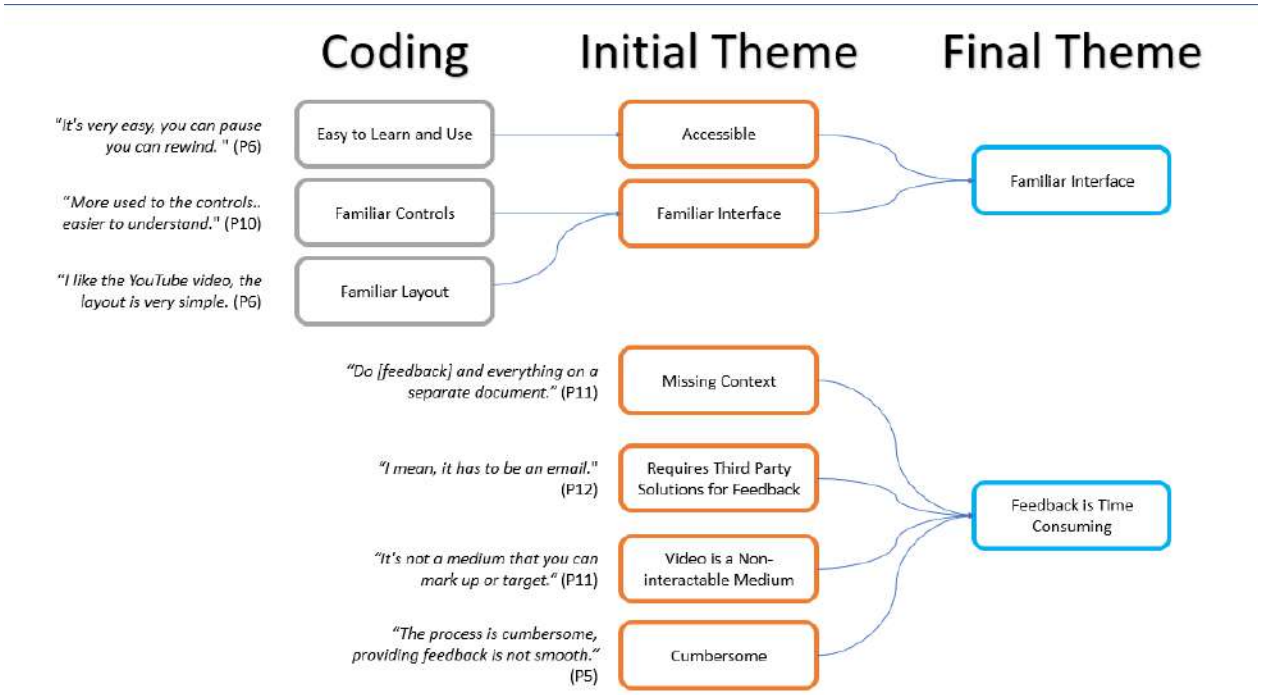


Figure 8.6: An extract of the thematic process conducted for the video assessment factor, showing the evolution of themes of *Familiar Interface* and *Feedback is Time Consuming* from cues, to initial themes and final themes.

8.3 Key Findings

From the thematic analysis in Section 8.2, four themes were selected as key findings. These themes are: *Audio and Visual Cues*, *Assisted Assessment and Feedback*, *Point of Observation during Assessment* and *Follow-up Dialogue Session*.

The *audio and visual cues* theme was selected because it emerged in both the framework and video assessment thematic maps. The *assisted assessment and feedback* theme was selected because of its prevalence and importance in the process of assessors conducting assessment and generating feedback. The *point of observation during assessment* theme was selected because of the contrasting themes for the framework and video assessment factors, which highlighted the importance of *complete control of observation* and the impact of assessment when *viewed from trainee perspective (gaze)*. And finally, *follow-up dialogue session* was selected because it emerged within all thematic maps.

8.3.1 Audio and Visual Cues

During the evaluation, a reoccurring discovery from participants was the importance of the audio and visual cues presented in the framework and video assessment factors. Findings from the analytical data factor showed that without direct observation of the trainee, it was impossible for participants to know the context of trainee actions. The prevalence of comments from participants which included mention of audio or visual cues, signified how important these visual and verbal cues were for participants being able to identify what the trainee is doing, and how they are coping with the tasks. As such, for conducting assessment, assessors considered it crucial that they can identify and assess the visual and audible information of the trainee in the simulation to understand the context of tasks relative to the overall trainee performances.

The importance of the audio and visual cues are easily identified when analysing Tables 8.3 and 8.2, which show how participants identified the confidence level of the trainee as part of their assessment. Participant ID's with asterisk (*) had no simulation audio during the evaluation. Out of the data-sets, when excluding the 2 participants who had no audio, in 17 out of the 22 observations participants used a combination of visual and audio cues to determine the trainee confidence level in executing the tasks. These results are comparable for industry and academic backgrounds, suggesting that audio and visual cues are a universally important element in the assessment process. Analytical data was not included in tables as no participant was able to estimate the confidence of the trainee due to the lack of audio and visual cues which provided no context to the trainee performance outcomes (see Section 3.8.1).

It is difficult to determine if those that utilised a combination of audio and visual cues were able to acquire a better understanding of the trainee performance, as participants clarified the traits of confidence differently, preventing comparisons of confidence estimations. However, what is certain, is that without context, the predefined outcomes of the analytical data had participants unsure about the assessment:

> *"Just that. How would you? I can't tell. You can't hear what they're thinking. We don't know how they [completed the task], how fast they went, or how much help they got. Did they request help? Did they do a double check? I can't tell anything" (Participant 10)*

Table 8.2: The cues of the observation from academic participants referred to during their judgement, assessment or feedback of the trainee. Participant ID's with asterisk (*) had no simulation audio during the evaluation.

Academic Observation Cues				
	Data-set A		Data-set B	
ID	Video Assessment	Framework	Video Assessment	Framework
1*	Visual	Visual	Visual	Visual
2	Visual	Visual	Visual	Visual
4	Visual, Audio	Visual, Audio	Visual, Audio	Audio
5*	Visual	Visual	Visual	Visual
6	Visual, Audio	Visual, Audio	Visual, Audio	Visual, Audio
10	Visual, Audio	Visual, Audio	Visual, Audio	Visual, Audio
12	Visual, Audio	Visual, Audio	Visual, Audio	Visual, Audio

Table 8.3: The cues of the observation from industry participants referred to during their judgement, assessment or feedback of the trainee.

Industry Observation Cues				
	Data-set A		Data-set B	
ID	Video Assessment	Framework	Video Assessment	Framework
3	Visual	Visual, Audio	Visual	Visual, Audio
7	Visual, Audio	Visual, Audio	Visual, Audio	Visual, Audio
8	Visual	Visual	Visual	Visual
9	Visual	Visual, Audio	Visual, Audio	Visual, Audio
11	Visual, Audio	Visual, Audio	Visual, Audio	Audio
13	Visual, Audio	Visual, Audio	Visual, Audio	Visual, Audio

Most importantly, after evaluating the factors, participants were asked to determine if the analytical data factor was suitable for authorising the trainee to operate in a real-world scenario. Given the scenario that analytical data output showed that the trainee had passed all outcomes, and completed objectives correctly, with the indicator 'Pass', would they as assessors, be satisfied with the trainee performing the same simulation scenario in a real-world environment:

> *"No. Its not just the outcomes we look for, its about the way [the trainee] works. That's what I look for in [classroom-baesd assessment]." (Participant 13)*

Not a single participant in the study would authorise the trainee based on the analytical data, with all sharing similar concerns with regards to lack of context to the trainee actions. However, participants were more accepting to both the video assessment and framework factor under certain conditions, which demonstrated over-time the trainee's continual application of knowledge in a variety of training simulations. The reason participants were open to the idea for framework and video assessment factors was because they were able to witness for themselves, the trainee conducting the assessment, visually observing them and determining if they have a grasp of the knowledge required. Without assessing visual and audio cues from the trainee performance, participants were not confident in the predefined learning outcomes of the analytical factor:

> *"I would want to witness them carrying out the actual activity." (Participant 11)*

> *"I don't think it's a good method because it doesn't give you a lot of context." (Participant 3)*

> *"Based on what we do just now [in classroom-based training], we physically see them and we observe them. It could be that [he] was just lucky guessing, I just don't know." (Participant 7)*

As discussed in Section 2.4, authentic assessment (Wood et al., 2013; Wiggins, 1990) focuses on a method of learning which is practical and natural for the real-world conditions in a professional setting, focusing on authenticity of the performance competency of the trainee and logical application of knowledge. Quotes from participants 3, 7 and 11 suggest why visual cues are crucial for the assessment of trainees in virtual reality training simulations. By directly witnessing the context of the activity, the participants are able to see the trainee logically apply their knowledge to resolve the problem. Without the visual and audio context, the outcomes of the trainee performance have no merit associated with them, as the predefined outcomes are not considered reliable indicators of trainee competency. As such, the outcomes from analytical data could have been the result of luck or guess work. This means that merit and trust can only be acquired by having assessors validate through direct visual observation of the trainee logically performing the tasks and achieving the objectives within the simulation. This finding suggests that without direct observation of the trainee, authentic assessment cannot be conducted, as true performance of the trainee is not evaluated, rather the predefined conditions of the simulation are assessed, detaching the assessment from the context of the trainee performance and application of knowledge.

Visual cues are more than just seeing the trainee conduct actions, it is understanding the motivations and thought-process the trainee is going through to apply the knowledge to a set of problems. For example, when using the video assessment factor, participants used the perspective and gaze of the trainee to determine the trainee thought process, allowing the participant to see through the trainees eyes (see Section 8.3.3). As a consequence, visual cues that inform the assessor of the trainees attention and thought process will improve their confidence of judgement in trainee performances. This suggests that additional features for eye-tracking of the trainee (see Section 9.3), similar to work by (Thanyadit, Punpongsanon, and Pong, 2020) would further improve assessors confidence in assessing and judging a trainee.

While visual cues provide the context to what the trainee does during the assessment, audio cues provide information for how the trainee is coping. The information from the audio cues are verbal comments made by the trainee which highlight potential problems or concerns of the trainee. Audio cues of the trainee performance enhanced the context of the observation for participants. The audio cues included in the trainee data-sets were inspired by observations of trainees in industry training centres (see Section 4.1), forming the requirement R5. Trainees were observed in assessments and training operations talking to themselves, hinting at potential methods of completing the task, hoping that the assessor would offer assistance to point them on the correct path. When participants were asked to reflect on the trainee confidence immediately after the using the factor, participants expressed how they primarily identified areas of low confidence in the trainee performance from the verbal comments made by the trainee:

> *"He was slow, but methodical, and he was sort of questioning himself." (Participant 8)*

> *"It doesn't sound particularly confident. I mean, he sounded like he didn't really know he was kind of having to figure things out as he went along." (Participant 13)*

> *"His vocal tone, especially when he was isolating it. He just guessed and he wasn't methodical about it." (Participant 7)*

> *"So they certainly seem to be asking themselves a lot of questions as well." (Participant 11)*

Participants almost always identified the audio cues of the trainee performance, with the above quotes all reflecting on the verbal comments of the trainee and how it impacted their judgement on the trainees confidence level and overall understanding of the trainees knowledge and approach for the simulation objectives. The tone and contents of the verbal audio cues provided context to the assessment of the trainee performance, as the participants could directly see and hear the trainee, allowing them to understand the contexts to why and what the trainee struggled with during the performance:

> *"I can hear the [trainee] talking through the things in his mind as if he wants to let the person know, his thought process so that he can justify it and talk about it later on. He wants to be reviewed." (Participant 12)*

The audio cues played a role in how participants provided feedback, with participants focusing their attention of feedback to target the areas of concern as expressed by the trainee. As such, the feedback was delivered to be supportive, re-enforcing or correcting gaps in trainee knowledge as was perceived from the audio cues. These situations show why the context for visual and audio cues are considered important when conducting assessment and providing feedback, as the cues provide additional information for assessors that enable them tailor the feedback to a trainees individual concern or problem.

When combined, audio and visual cues provide a strong contextual reference to both the actions conducted by the trainee and their thought process when performing actions. Information from these cues provides a reliable platform which presents assessors with in-depth context and knowledge of the trainee performance, enabling assessors to determine if trainees are competently performing the objectives:

> *"You are seeing that somebody is competent enough, so the video is a good place to start. Then you can see whether someone approaches [the tasks] the right way." (Participant 9)*

During the evaluation, participants relied on these cues to ensure that the objectives were logically and methodically completed. As discussed previously, without context to learning outcomes, participants were not confident in providing assessment or feedback, or measuring the competency of confidence of the trainee. These examples clarify the importance of audio and visual cues for authentic assessment.

In summary, if virtual reality is to supplement or replace traditional real-world training, it is important that the assessment and feedback incorporates audio and visual cues to enable a direct observation of the trainee performance, allowing the full impressions of trainee to be presented to assessors, as found in real-world assessment (Fardinpour, Reiners, and Wood, 2018). These findings underpin the importance of the audio and visual cues in assessors judgement of competence and enables authentic assessment and contextual feedback to target the individual concerns of a trainee, as acquired from the nuances in their performance presented by the cues.

8.3.2 Assisted Assessment and Feedback

The second key finding from this evaluation showed the importance of embedded authentic assessment and contextual feedback tools within the framework. During the evaluation participants only generated feedback using the framework factor, with the process of providing feedback described for analytical data and video assessment factors. For the purposes of this evaluation it was clarified that the trainee was located in the other-side of the world, therefore feedback had to be conducted remotely. This clarification was added to stop participants from attempting to use face-to-face feedback they are accustomed to from classroom-based assessment.

Embedding assisted authentic assessment and contextual feedback tools into a single framework, streamlined the process of conducting assessment. The framework enabled an all-in-one method for observing the trainee for assessment, followed by quick generation of contextual feedback within a single assessment environment. The framework included a variety of assisted assessment tools and features which helped participants judge trainee performances similar to classroom-based assessment.

Assisted Assessment

Assisted assessment refers to the inclusion of features and tools that help participants conduct assessment, making their role easier, and providing them additional information that enables them to reliably and confidently conduct assessment of a trainee performance. The summative information (see Section 6.5.1) and augmented timeline (see Section 6.5.2) features were shown to improve participants ability to conduct in-depth assessment, which in turn, provided participants with confidence that they had correctly judged the competency of a trainee.

During the evaluation, 3 participants who had used the framework first, followed by the video assessment factor, skipped sections of the video assessment factor, stopping intermittently to clarify the progress of the trainee. With prior knowledge from the initial use of the framework factor, these 3 participants knew where trainee mistakes where during their performance. As such, the 3 participants monitored the simulation environment and the positioning of the trainee to determine what stage the trainee was currently at in their performance. These participants stopped skipping when they anticipated the trainee was about to conduct a failure or mistake. In doing so, participants demonstrated that given prior information of the trainee performance, participants were confident in skipping sections of the assessment and focusing on the areas of trainee performance that were known problems or concerns. In contrast, participants who used video assessment first, were forced to watch the entire observation, as they had no prior familiarisation with the trainee performance. Putting aside concerns regarding the validity of assessment when skipping data, these findings identify that prior information of the trainee performance can assist the participants identify the most relevant aspects of a trainee performance.

Participants who skipped sections of the video assessment in the data-set A, only stopped at the areas of interest when they knew the trainee was going to complete an objective or would ask a question. However, as participants were not informed data-set B was different between the framework and video assessment factors, participants skipped ahead, but stopped when they noticed their prior knowledge was inconsistent with the trainee performance. Once participants realised the data-set was either different, or areas of the performance were unfamiliar, participants stopped skipping. When asked the motivation behind skipping, participants all indicated they were attempting to look for areas of the performance where a trainee was completing objectives or experiencing difficulty:

> *"I'm looking to try and find the key points in the assessment where they would carry out a certain activity [...] or they would carry out a particular procedure. You might jump to the key points. And then if you're unsure, go back and watch that action in more detail." (Participant 11)*

> *"I was trying to get to the key bits where things happen." (Participant 13)*

> *"I'm checking if the stools there or not. So that that I can know if the trainee did pick it up. Okay, so I said [to myself] if [the stool] still going to be here in five seconds [the*

> *task] is done, if [the stool] does not go away in five seconds, then [the task] is not." (Participant 5)*

With prior knowledge of the trainee performance, participants would skip sections of the video assessment and focus on the events where the trainee was completing actions. As noted by Participant 5, knowing how the simulation outcomes impact the content within the virtual environment, participants skipped ahead looking to see when elements of the virtual environment were different. In doing so, participants would ignore all other information provided by the video, focusing on the modification of these changes to identify what stage the trainee was at and to determine if the task was completed. However, it is important to consider, that without prior knowledge, skipping these sections is likely to result in loss of data from the trainee performance:

> *"You'll overlook something [...] You would miss [trainee problems] if you didn't go back and watch it. So I guess in that sort of mentality, you would have to watch the whole thing from start to finish." (Participant 11)*

When questioned on the matter, Participant 11 acknowledged this concern, clarifying that they would have had to watch the entire trainee performance if they had not already known where the failures where. This also aligns with the findings from participants 3, 6, 9 and 12 who operated video assessment as the first visual factor during the evaluation, all of whom watched the full video without skipping.

These findings highlight the importance of embedded assisted assessment information into the observation. As the framework provided this information through the summative performance indicators and timeline, the framework was best suited to meeting the needs of participants. Moreover, with the prior knowledge of the trainee performance offered by the framework, participants were best prepared to identify and spot concerns on the lead-up to a known fault within the simulation, rather than waiting until after the event, as was the case with participants skipping sections with the video assessment factor, as discussed by Participant 11.

However, the assisted and assessment tools should not be limited to what has been presented in the framework. Other features for transcribing trainee audio data as conducted by Raij and Lok (2008), offer valuable opportunities to improve the overall method of improving the assessment process for assessors. As demonstrated by the framework, it is possible to incorporate the assisted assessment information captured from trainee performance information into direct observation methods of assessment. Overlaid information from the training simulation can provide the assessor with a reliable reference to how the trainee has performed in predefined objective tasks. With this information, the observation of the trainee can be focused on and in the lead-up to known areas of interest within the simulation. Furthermore, these findings corroborate answers received from assessors in the requirement collection study conducted in Section 4.1, which signified the importance of action identification for trainee actions throughout a trainee performance. As such, these findings justify the framework design and implementation for assessment, which embedded assessment aids that provide identification of actions and events within the trainee performance using the summative performance indicators (see Section 6.5.1) and augmented timeline (see Section 6.5.2) tools and features.

From the results presented in this section it can be concluded that assisted assessment improved the usefulness and usability of conducting assessment, with the augmented timeline and summative performance indicators used to identify the areas of immediate concern, lessening the cognitive overhead for the assessor when using the framework. These assisted features could be further enhanced with transcribed audio logs of the trainee (Raij and Lok, 2008) to ensure that context remains for sections that are skipped by the assessor.

Embedded Feedback Tools

Embedded feedback tools included the video narration tool (see Section 6.6.1), evidence capture tool (see Section 6.6.2), and drawing tool (see Section 6.6.3). Paired with the assessment tools, the authentic assessment and contextual feedback process was streamlined, making the participants job easier (Hanoun and Nahavandi, 2018) and enabling feedback to be contextual and targeted.

The video narration tool was the primary method of generating feedback for the trainee when using the framework. Participants highlighted that the inclusion of the feedback tools within the assessment environment enabled them to generate contextual feedback in a time efficient manner. In comparison the video factor, provided no embedded method of providing feedback, and as a consequence, participants were dismayed at the use of external feedback tools which both reduced the efficiency of conducting assessment and limited the context and content of feedback. When asked how participants would provide feedback of virtual training simulations, the general consensus was that participants would copy the framework approach, as they considered both the process and content of the feedback very effective:

> *"I think seeing [the framework] for virtual reality assessment, it's certainly gives me confidence that in the future, that's something that we're going to be able to do." (Participant 11)*

Written notes in combination with video timestamps was the main approach participants utilised for generating feedback for the video factor. Participants discussed that for feedback to be generated, a word document would be used on a second monitor, allowing notes to be written while the performance is being observed on the primary monitor. When probed further how these written notes would be structured and how the trainee would be able to understand the content and context of the data, participants relied on supplying the YouTube video with timestamps attached to the relevant sections of the written notes. For example, if feedback was generated at timestamp 2:03 in a video, the participant would mark that time on the document, and write the feedback for that section, limited to that fault. In explaining on this process, participants reflected on the challenge of providing feedback via this method, as it would be time consuming and detached from the assessment environment:

> *"The process is cumbersome for providing feedback. It's not smooth, and you can't really identify clearly, what went wrong without a lot of jumping through hoops or using [Microsoft] teams, using email, using phone calls. And it's easy to see that the information that you generate for [the trainee] about the failures could get misconstrued or lost." (Participant 5)*

As such, providing feedback external to the assessment environment is less efficient and effective and impacts the cognitive load for the assessor, as their attention becomes diverted between multiple applications, potentially causing the supplied feedback to be misconstrued. This is avoided by the framework as it combines conducting assessment and generating feedback within a single framework environment. This contextualises the trainee performance and using the assisted assessment tools and features, participants are able to provide targeted feedback:

> *"For example, when I watched the second [data-set B] using video, I wasn't entirely sure whether [the trainee] completely cleared the area or not [...], whereas with the [framework] [...] you've got the [timeline], then you know, you can see whether all these things have been done or not. [The trainee information] is just there for you." (Participant 1)*

However for the video factor, identifying and generating feedback relied on the attention, effort and technical skill of the assessor using multiple applications, such as MS teams or email, to generate feedback that is contextual and provided in a time efficient manner. Generating a video narration externally from the framework requires significant technical and video editing knowledge, which would not be viable for the targeted end-users, who struggled with adapting to the interface design of the framework, which was inspired by video editing suites (see Section 6.5).

These are significant findings, as they show that embedding feedback into the same environment as assessment improves the experience and streamlines the process. As Jonsson (2013) note, if feedback is delayed for too long, it becomes redundant, therefore, by streamlining the feedback process, assessors are able to quickly and effectively generate feedback within a single assessment environment, as demonstrated by the framework.

The inclusion of the drawing tools within the framework also proved useful for participants, as they were able to highlight the exact focus of their feedback. Inspired by the study in Section 4.1, the tool was designed to mitigate the loss of human gestures that were common in classroom-based assessment. When generating feedback, participants used the drawing tool in combination with the video narration tool to generate contextual feedback which guided the trainee, informing them where they should focus their gaze when analysing the feedback (Valenzeno, Alibali, and Klatzky, 2003).

While participants often appeared satisfied with the outputted feedback from the framework factor, 3 participants indicated they would still prefer written notes to accompany the video narration feedback. However, participants were unable to clarify their reasoning behind this, mostly arguing that it is best to provide feedback that satisfies all of the senses and makes the feedback easily accessible to the trainee. Interestingly, these comments were only present from participants with an academic background, rather than industrial training, potentially suggesting that participants were sticking with what the know best, influenced by how they currently educate:

> *"I think I would supply a written note, because it's always good to use, audio, visual and in writing [...] give [the trainee] everything to use all the senses." (Participant 10)*

However, a concern formed when analysing the written feedback for the video factor. Participants indicate that written feedback could dilute the feedback quality, making the contents more generic and less targeted to an individual trainee:

> *"For simplicity, because probably for a lot of them or many of them, they may have failed in the same sections. So I can copy paste the feedback from one to another." (Participant 6)*

These results suggest that the inclusion of written feedback was requested as it was easier and quicker to provide through repetition. However academic participants approach of repeating the same feedback to multiple participants is inadvisable for authentic assessment, as the feedback is not targeted to the individual concerns or faults of a trainee, reducing its effectiveness (Jonsson, 2013). So while, written notes could be an additional method of conducting feedback, it is important that written feedback is not the sole source of providing feedback (see Section 2.8.1).

From this discussion it can be concluded that the combination of assisted authentic assessment and contextual feedback tools within a single framework environment improve the efficiency and effectiveness of the assessment process. Assisted assessment tools provide additional context and can help identify the areas of immediate concern within trainee performances. Embedded feedback tools demonstrated effective generation of contextual and targeted feedback that assessors consider as useful and usable to support the learning of trainees. The combination of the authentic assessment and contextual feedback tools streamlined authentic assessment operations, lessening the cognitive overhead for the assessor by providing them with the tools necessary to conduct assessment and generate feedback within a single framework environment.

8.3.3 Point of observation during Assessment

Contrasting themes for the framework and video factors emerged regarding the perspective of observation during assessment. For the framework, the theme *complete control of observation* emerged (see Figure 8.1), presenting the unrestricted approach for observation and assessment, which enabled participants to view the trainee performance from any perspective, providing confidence in their judgement of trainee competency. In contrast, the *viewed from trainee perspective (gaze)* theme emerged from the video factor thematic map (see Figure 8.5), identifying the usefulness and problems of observing and assessing from the perspective of the trainee, which provided additional context, but caused 5 participants to experience motion-sickness (Jerald, 2016).

For usefulness, the video factor provided the ability for participants to see the training performance through the eyes of the trainee:

> *"When you can see the vision through his eyes, I could see what he sees [...] You know, he was looking at lever he was looking at the gauge, he was looking at the lever. So I think my confidence [in the trainee] went up a bit when I saw that." (Participant 10)*

This perspective provided participants with a potential opportunity of being able to understand the trainees thought process when conducting tasks and objectives, letting them read the situation and understand the logic and context behind the actions of the trainee:

> *"With the video, you only see things that seem important to [the trainee]. So you're already getting dragged along a little bit with them. So in a way, it makes it easier to follow along with the video, because you're already seeing the choices that are made by the [trainee], and see the situation that is running through their mind." (Participant 12)*

Using the eye-gaze of the trainee, participants were able to identify what the trainee was looking in the simulation environment to determine their thought process. By understanding the perspective of the trainee, participants are guided in their assessment process, improving their ability to relate the trainee. However, viewing the performance from this perspective induced problems, causing discomfort in 5 of the 13 participants:

> *"I was just getting a little bit dizzy as [the trainee head] was moving around [...] the movements in the video are very erratic, because you're looking at it from the [trainee] perspective." (Participant 4)*

For others, the perspective of observation started inducing signs of motion sickness, due to motion of the trainee view altering their perspective:

> *"Was a little bit [dizzy] like, I mean, I wouldn't to watch a few of them I'll say that [...] just the head like the pointing everywhere, it's not a single point of view." (Participant 5)*

In comparison, the free camera tool in the framework, provided participants with unrestricted control of the observation (see Section 6.3.2). Because the video factor was predefined, the assessment and observation was restricted to being observed from the perspective captured from the trainee's gaze. While multiple static perspectives could be used to overcome this issue (Thurston and Martin, 2011), long term assessment or observation from the perspective of the trainee has implications for the health and well-being of participants:

> *"When you're looking at the [trainee perspective] for more than say 20 minutes, I think I could get quite sick." (Participant 7)*

With several participants indicating motion sickness or discomfort, the perspective of viewing the trainee from their perspective is considered unsuitable for long-term use as the primary method of observation. However, participants were impressed enough with the information acquired from viewing the trainee perspective that they expressed including a variation of the perspective as an additional assessment aid for the framework:

> *"I think with the [framework] an improvement would be so you can also watch it from the trainee's perspective [like the video factor]. If you could do it from both, that would be more valuable. Like if you can use two screens or something, you know, you could have [the trainee's perspective] and at the same time using the [free camera tool], a small picture and picture type thing. That would be really good." (Participant 13)*

As such, the trainee perspective was found to be useful as an additional assessment aid, rather than being the primary method of observation, as was the case with the video factor. Operating as a picture-in-picture (PiP) tool, the perspective of the trainee can be used to enhance the information available to the assessor. However, this PiP implementation may still cause discomfort or motion sickness. Alternatively, the use of eye-tracking equipment may be best situated alleviate this issue (Meißner et al., 2017), allowing the acquisition of the trainee gaze, providing the context and information to the assessor. This would be classified as an additional feature within the assisted visualisation of the trainee tools (see Section 6.5.3).

8.3.4 Follow-up Dialogue Session

The final key finding from this evaluation was the prevalence of demand for a follow-up session with the trainee which would involve the assessor and trainee reviewing the assessment together to discuss the learning outcomes, and provide the trainee with the opportunity to directly question the assessor. These themes were prevalent in both academic and industry responses, indicating that the question and answer session is an important element within learning, potentially because it forces learners to review the feedback they receive in preparation. As stated by Jonsson (2013), if learners do not view or utilise the feedback, they do not learn from their mistakes. Therefore, by following up on the trainee, the assessor can determine if the trainee has understood the feedback, and check if they need any further help or support. While Hanoun and Nahavandi (2018) argues that performance must be reviewed and acted upon immediately after an assessment, this is not always possible for remote forms of assessment. Time-zone differences or commercial considerations of companies, it is probable there will be a delay between the completion of an assessment by a trainee, and the feedback provided by a assessor. As a consequence, for after-action assessment, it is best to focus on the process of streamlining the assessment and generation of feedback, which can then be followed-up with a dialogue session.

Follow-up dialogue sessions varied based on the factor used. For the few participants who attempted to provide feedback with analytical data, they relied on a follow-up dialogue session to question the trainee on their performance, hoping to acquire information that would provide context to the predefined outcomes. Alternatively, for the video and the framework factors, participants considered the follow-up feedback as an opportunity to further review and discuss areas of the trainee performance which required feedback to be generated. For the framework and video factor, participants envisioned using the factors as a backdrop to the discussions, allowing visual references of the simulation to guide the topics of conversation. This would provide the opportunity for the trainee to ask the assessor questions regarding their training performance, similar to the work of Stone, Snell, and Cooke (2016). Participants considered these follow-up sessions as a method to validate that the trainee has absorbed the feedback initially supplied to them.

The study conducted in Section 4.1 highlighted that in classroom-based feedback, assessors are able to use facial cues of the trainee to determine if they have absorbed and understood the feedback. As such, there was a concern that when generating feedback remotely, the trainee may not be able understand the feedback. To mitigate this concern, participants considered a follow-up dialogue session as important to allowing them to check on the trainee and see how they are coping, and determine if any further assistance is required:

> *"Follow up I guess, with a team call, but I don't know how appropriate that is with the other side of the world, you know, but certainly [the trainee] needs to be sent the [feedback] information and have an understanding of what that that problem is, so you want to build the feedback first. But, you also want to know that they can accept and can understand what it is, and what it the information that you're trying to give them. I think that's the element, you lose with an email, you can send a lot of good information via email, but you lose that connection to say "Did you understand? Oh, no, you didn't? Why don't you understand?", you know, that two way conversation." (Participant 9)*

With the feedback supplied for the video factor conducted via email, participants expressed the importance of feedback being useful and usable by the trainee, allowing the content to be digested. Because it is difficult to determine the trainee response to feedback remotely, follow-up dialogue sessions were seen as an opportunity to enable one-on-one discussion between trainees and assessors. However, follow-up dialogue sessions were mostly driven by situations where the participant felt the generated feedback was poor (analytical data and video assessment factors), or the feedback alone would be insufficient for correcting major errors, such as critical safety failures (see Section 3.8.1) in a trainee performance:

> *"Maybe on this particular example, given the severity of the failures, it might be good to have that face to face, because part of feedback is also gauging their reaction to that feedback and understanding of the feedback that you've given." (Participant 11)*

For example, a follow-up dialogue session was deemed urgent for data-set B, which contained several mistakes and one critical safety error, as participant 11 felt that forceful feedback was required as the trainee demonstrated a severe lack of knowledge. However, the need follow-up dialogue sessions also depended on the quality of feedback provided by the factor used:

> *"For me providing feedback, it would be [the framework], because I can actually provide real feedback." (Participant 11)*

As the feedback supplied with the framework was considered equivalent to that provided in traditional classroom-based environments, which included context and targeted feedback, there was less demand for follow-up dialogue sessions. However, it was still considered a potentially useful education and support tool for instances where trainees needed further prompting:

> *"I'd ask him why he made that decision, he might not be able to tell me. But I may ask him [...] just to see what he says." (Participant 7)*

Follow-up dialogue session were not restricted to discussing feedback, with participants envisioning it could be used to perform a live walk-though of the performance, allowing real-time analysis and discussion among individuals or groups:

> *"I rather imagine, though, that we could both [trainee and assessor] be in this session together, like, you and I are just now to be in the same environment with [the trainee]. With maybe one camera, and we could do what we're doing now and could review the playback and see their mistakes, and it would give us a chance to chat about it." (Participant 12)*

With context, the follow-up dialogue session allows both the assessor and trainee to review the simulation performance and discuss it together, reproducing the after-action discussions and evaluation that occurs during classroom-based training:

> *"I think you can also ask what he was thinking. Like, you know, [the trainee] could say he was confused for a moment, but he just wanted to make sure, or I [as the trainee] was considering all the implications." (Participant 5)*

While the follow-up dialogue session is dependant on the performance of the trainee and the factor used to generate feedback, the concept and implementation enables trainees to be supported remotely if necessary for additional issues that are not identified. As such, follow-up dialogue sessions are an important tool that can be used to support trainees remotely. These findings match the assessment process conducted by Stone, Snell, and Cooke (2016), and share some similarities with the live streaming component (Howie and Gilardi, 2020) (see Section 6.5.4), suggesting that real-time collaborative authentic assessment and contextual feedback could be beneficial.

8.4 Evaluation of Framework Tools

The framework components for observation (see Section 6.3), assessment (see Section 6.5) and feedback (see Section 6.6) included several tools and features that were designed to assist the assessor (Hanoun and Nahavandi, 2018). These tools and features were evaluated by participants who discussed their use and influence on the assessment and feedback process.

8.4.1 Observation

To provide direct visual observation of the trainee, the framework embedded a *Free Camera Tool* and *Replay Features*, which enabled participants to view the trainee performance from any perspective at any time period.

Free Camera Tool The free camera enabled participants to perform unrestricted observations of trainee performance. Participants often praised the concept, but criticised its difficulty to use (see Section 8.5.3). The free camera was considered an important asset in participants authentic assessment and contextual feedback of the trainee, as the feature ensured that the participants perspective monitoring the the trainee performance was not obscured:

> *"This, I think is like an improved version of the video because you have a better view of the trainee [...]. When you are reviewing and [assessing] trainees, I think especially for feedback, I feel [the framework] is the best method." (Participant 1)*

The camera was also found to be influential in participants response to providing feedback using the video narration tool, with the ability to alter the perspective providing additional context that supplemented the verbal narration content to provide targeted feedback. Participants were observed during the evaluation replicating their classroom-based protocols for monitoring trainee performances, by exploring the environment and monitoring the trainee from different perspective.

And, because the simulation was after-action, participants were able to view the intricacies of trainee performance without intervening or being intrusive to the trainee, a concern raised from classroom-based training (see Section 4.1).

These findings show that the free camera tool was effective for assessors, allowing observation without any restrictions, obstructions or concerns that their observation would impact the trainee. As such, the free camera tool (see Section 6.3.2) met the requirements R1, R2, R5 and R6 from Section 4.2 and successfully complemented the other observation, authentic assessment and contextual feedback tools and features included within the framework

Replay Features Participants found the replay features useful and effective, allowing reliable observation and assessment of the trainee which provided participants with confidence that their judgement of the trainee was reliable. For participants, especially those who used the framework as the first factor in the evaluation (see Table 8.1), the replay feature was used to rewind and replay sections of the trainee performance that was initially missed due to poor positioning of the free camera tool:

> *"It's easy to go back and forward to find particular sections if you want to review those. And the [timeline] time markers and time-stamps were good. Plus, it showed me the data set, showing me the pass and fail outcomes." (Participant 10)*

These results indicate that the replay features met the requirements of R1, R2 and R4 from the background study conducted with assessors in Section 4.2). The replay features provided the ability for assessment to be replayed, or skipped as required (see Section 6.3.3), overcoming concerns discovered with classroom-based observation and assessment (see Section 4.1).

8.4.2 Assessment Tools

To provide authentic assessment, the framework embedded several assessment tools and features that were not present in analytical data and video factors. These tools included the *Summative Performance Indicators*, *Augmented Timeline* and the *Assisted Visualisation of the Trainee Tools* which provided enhanced abilities that improved the assessment capabilities of assessors.

Summative Performance Indicators The summative performance indicators provided participants with the means of quickly familiarising themselves with the trainee performance. With the summative performance indicators presenting the outcomes of the trainee performance in definitive pass or fail outcomes, academic participants - who were unfamiliar with the oil and gas training procedures- were able to identify failures in trainee performance. As such, the summative performance indicators were seen as an additional assessment aid that provided participants with knowledge of the trainee performance prior to beginning the assessment:

> *"I mean, this is a good addition to the video, but I don't think on its own its okay [...]*

I don't think it's enough to assess if someone's competent to do the job or not [...] it could have been good luck as much as anything else." (Participant 9)

Participants stressed that alone the summative performance indicators would be inadequate, and required context to through direct visual observation to conduct assessment and generate feedback. These findings show that the summative performance indicators (see Section 6.5.1) successfully met the requirement R3 from Section 4.2 for assessors.

Augmented Timeline The augmented timeline was found to improve the assessment process, with participants using the additional information to influence their assessment of the trainee performance, enabling deeper insight into the performance of the trainee. The augmented timeline provided in-depth context to the actions of the trainee, displaying key-points of data in sync with the direct visual observation of the trainee, allowing easy identification for the trainees progress within the simulation:

"There was a panel running along horizontal at the bottom is when something in his hand. You can see and also there was the task that he passed and failed running up vertically, all of that is missing [with video]. That was a detriment, because [with it] you're able to see [what stage the trainee is at] quickly, even if you've lost track [...] you can quickly catch up so you know, what stage of the assessment you're at." (Participant 10)

Participants often contrasted the differences between the video and the framework factors, highlighting the importance of the assisted assessment features provided by the augmented timeline in the framework for keeping track of the trainee progress. These findings show that even with minimal exposure to the framework, participants reflected on the importance of the augmented timeline, showing that the augmented timeline left a definitive impression, and improved the process for conducting assessment. As such, the augmented timeline (see Section 6.5.2) met assessors requirements of R1, R2 and R4 which were identified in Section 4.2.

Assisted Visualisation of the Trainee Tools Assisted visualisation of the trainee tools included a skeleton structure visualisation to identify the posture and movement of the trainee, along with the audio cues captured from the microphone of the trainee. As already been discussed in Section 8.3.1, audio cues were a key finding of this evaluation, signifying the importance of audio cues in providing additional information that can contextualise the trainee performance for the assessor.

For the skeleton structure visualisation, participants considered it unnecessary, with most not using it. The simulation already contained a virtual avatar, which was found to provide a sufficient representation of the trainee, which enabled participants to identify the body posture and movements of the trainee. Participants that did use the tool, used it to see through the trainee avatar without having to move the free camera tool. As such, the skeleton structure visualisation was not deemed as necessary for assessment of the simulation scenario used for the evaluation:

"I don't really know [VR] so I'm guessing the skeleton won't be that important, or I didn't feel the need for it. There's no right or wrong answer. It's just, I don't know what to feel." (Participant 4)

However, these results are only relevant for the evaluation scenario, as other training simulation scenarios may require actions and objectives to be completed using specific posture (Bernard et al., 2018), in which-case if the virtual avatar is inadequate or modified for the comfort of the trainee (Jerald, 2016), the skeleton structure visualisation would be relevant. These findings indicate that the assisted visualisation of the trainee tools (see Section 6.5.3) met the requirements R1, R2 and R5 from Section 4.2).

8.4.3 Feedback Tools

Feedback tools were embedded into the framework that enabled effective, useful and usable feedback to be generated. These tools include the *Video Narration Tool*, *Evidence Capture Tool* and *Drawing Tool*, which when combined with the assessment tools (see Section 8.4.2), streamlined the assessment process for the assessor.

Video Narration Tool The video narration tool was often viewed as the best method of providing feedback in virtual training simulations, with many participants praising its implementation, along with the feedback it generated. Participants reflected that the feedback supplied using this tool was far beyond their expectations, referring to existing methods of feedback as inadequate. When compared to the narrated video, other factors which required external solutions where time consuming and considered less effective:

"This is a more comprehensive way of [providing feedback], and ultimately, it is more time efficient." (Participant 9)

Furthermore, because the feedback included context of the trainee performance to supplement the narration by an assessor, the feedback was considered immediately effective, with participants noting that they anticipate trainees would be more than satisfied with the feedback. Five participants were so confident in the context and content of the feedback supplied by the tool that they stated that they would require no other supplemental or additional feedback, such as a follow-up dialogue session (see Section 8.3.4), since the tool satisfied all their requirements with no perceived improvements possible. When asking how the generated feedback compared to classroom-based feedback Participant 9 commented:

"Pretty amazing, really, if that's a way of giving feedback over what we did previously, which was just, you know, me talking to maybe send them an email or something, it seems quite archaic and I get this kind of system. It's pretty amazing way of giving feedback. And if you're talking about doing it on the other side of the world, you're

losing that face to face thing, but I think you're able to explain things so much more clearly." (Participant 9)

These findings imply that the video narration tool satisfies requirements R6 and R7 from from the background study conducted with assessors in Section 4.2, with design of the tool (see Section 6.6.1) proving successful to achieve comprehensive feedback that is comparable with classroom-based feedback.

Evidence Capture Tool The evidence capture tool was restricted to capturing only an image of the trainee performance to match classroom-based usage, for the requirement R7 (see Sections 4.2 and 6.6.2). In comparison, the video narration tool captured a video of the trainee performance which included additional audio cues, providing a contextual record of a trainee performance. As a consequence, the evidence capture tool was not considered useful as a method for providing feedback or storing contextual evidence of the trainee performance. Similarly, because participants in the study were not tasked with generating evidence for archival, participants did not see an immediate role or for the evidence capture tool (see Section 4.3.3). Participants occasionally captured a picture of trainee faults after generating narrative feedback, but did not elaborate further on their impressions of the tool, as their attention for feedback was focused primarily on the video narration and drawing tools.

As the evidence capture tool was not evaluated for its intended purpose, it is difficult to determine if the tool is considered useful. However, because it is known participants may stick to what they know and reject change once the are familiar with something (Mueller, Melwani, and Goncalo, 2012), participants positive impressions of the video narration tool likely caused the evidence capture tool to be considered a step back, with less features, context and content. As such, the evidence capture tool may be redundant in virtual forms of assessment, with participants viewing video narration tool as the best method for capturing content for evidence within virtual training simulations.

Drawing Tool The drawing tool was considered an important aspect of the feedback process, with participants discussing its positive impact on their ability to focus their feedback. As found in the literature (Valenzeno, Alibali, and Klatzky, 2003), participants used the drawing tool to target and focus their feedback to specific areas when capturing video for feedback (see Section 6.6.3). Working with the free camera tool, participants would position the camera in an advantageous position, start the video narration recording and draw onto the screen while verbally providing feedback over the trainee performance:

"The amount of detail you can get in the video capture is a brilliant feature [...], the drawing is very, very useful. [...] I would say that ultimately, all of that builds to create in a very comprehensive look at what the trainee does, and where they go wrong. And being able to feed that back in very accurately and detailed [...] it's a very, very comprehensive, and an excellent piece of equipment." (Participant 5)

The drawing tool enabled the feedback generated from video narration tool to be more effective and comprehensive. Video context, verbal narration, and drawings helped focus and highlight target areas during feedback for the trainee. These results show that the drawing tool was an effective method that extended the capabilities of assessment and was a suitable replacement for gestures, meeting requirements of R6 and R8 from Section 4.2, enabling feedback to be richer in detail, improving its overall quality and usefulness.

Participants ranked their order of preference for the three factors used during the evaluation to conduct authentic assessment and generate feedback for VR simulations. 11 out of 13 participants ranked the framework first, 12 out of 13 ranked video second, and analytical data (12 out of 13) was ranked third. Results from the analysis conclude that the overwhelming preference of the framework as the preferred method was due to the tools provided. Because these tools were embedded into the framework, participants praised the ability to quickly and efficiently generate feedback, without diverting their attention to other applications or losing track of their assessment progress. Furthermore, the assisted assessment tools and features enabled in-depth assessment of trainee performances, which allowed participants to engage with the data to analyse the augmented timeline and summative performance information in sync with their visual observation. These findings justify the approach used to conduct the research, highlighting that the existing research which influenced the design of the framework components (Hanoun and Nahavandi, 2018; Stone, Snell, and Cooke, 2016; Raij and Lok, 2008), has been validated in practice.

8.5 Limitations of the Framework

Although almost all participants selected the framework as their first choice for conducting authentic assessment in VR in the future, three limitations were identified: *network lag*, *technical interface* and *control difficulties*.

8.5.1 Network Lag

Due to the Covid19 pandemic, the evaluation was conducted remotely, streaming the framework (and other factors) from the author's PC using MS Teams (see Section 3.8), which introduced network lag. As the the free camera tool (see Section 6.3.2) required engagement from participants alter the perspective of observation, the network lag impacted the responsiveness of the framework:

> *"I know it's difficult to control remotely and it's much more so because it was lagging just now." (Participant 6)*

The framework was not designed to be controlled remotely, and was adapted to accommodate the COVID19 restrictions. These issues would not have been present in a face-to-face study.

8.5.2 Technical Interface

The framework included a variety of features and tools that were designed improve the assessment capabilities for participants, including the free camera tool (Section 6.3.2), replay features (Section 6.3.3), summative performance indicators (Section 6.5.1), augmented timeline (Section6.5.2), assisted visualisation of the trainee tools (Section 6.5.3), video narration tool (Section 6.6.1), evidence capture tool (Section 6.6.2) and drawing tool (Section 6.6.3). These tools were presented as part of the interface for the framework. However, when initially presenting the framework and tools available during the tutorial phase, participants appeared to become overwhelmed with the interface of the framework and struggled to absorb all of the information presented:

> *"I was kind of we bit distracted trying to use the tool. Maybe once you get used to it, you would stop looking at it [and thinking] Oh, that looks like a [UI] object" (Participant 10)*

Initial impressions of the interface show that participants considered the layout as too technical and unfamiliar, identifying a potential accessibility concern. Participants noted that the five minute tutorial was too short, and they would have preferred further time to learn the interface and comprehend how the tools and features could be used before beginning the evaluation. However, participants did become more comfortable as the evaluation went on, suggesting that lack of familiarity with the framework:

> *"It just feels like at the moment, I'm pretty clumsy with it with the controls. But I guess that's something that you would pick up and practice" (Participant 9)*

Given further time to familiarise themselves with the framework, these reactions would likely to have been lessened. However, due to the time limit for the evaluation, it was not possible provide a longer tutorial session, causing initial trepidation for the technical representation of the framework.

8.5.3 Control Difficulty

A main concern that was likely exacerbated by *network lag* and *technical interface* concerns, was the emergence of issues related to control difficulty for the free camera tool, which utilised a combination of the keyboard and mouse input from the participant to alter the perspective of observation (see Section 6.3.2). However, participants indicated they were unfamiliar with this control layout and as a consequence, participants initially struggled to operate the free camera tool:

> *"I think just the physical control with the keyboard would be the biggest challenge for most people." (Participant 11)*

Participants indicated that these control difficulties could be mitigated by a longer tutorial session that provided additional time to practice and by conducting the evaluation in-person to avoid the input delay caused by network lag:

> *"I think you should try and try and do some [framework evaluations] where you're not doing over a network, [rather] on the machine so they don't have [network lag]. I think it's giving is giving an unfair picture of what the [framework] could do." (Participant 1)*

Observations of participants interacting with the framework did show improved performance as the evaluation went on, with participants confirming that their confidence did improve as they become more familiar with the controls.

8.6 Chapter Summary

This chapter answered RQ6 and RQ7 and presented the findings from the evaluation of the framework, analytical data and video assessment factors with 13 participants from industry (6) and academia (7).

RQ6: *Are the tools and features provided in the framework suitable for conducting authentic assessment and generating contextual feedback of room-scale virtual reality training simulations?*

The findings from the evaluation answered RQ6 by presenting evidence the tool and features provided in the framework were suitable for assessing trainees in room-scale virtual reality training simulations. The embedded tools and functionality demonstrated several advantages and improvements over the other factors considered in the evaluation, with only minimal usability problems which can be addressed by improving familiarisation with the interface and longer tutorials to learn controls. These results validate that the framework presented in this work is not only a viable option, but an improvement upon existing solutions:

> *"It's much easier to analyse the data as a complete package with [the framework]" (Participant 9)*

Assisted assessment features of the framework were discovered to improve the assessment process, providing easier identification of trainee actions and events within the simulation performance, with summative performance indicators quickly showing pass or fail conditions for simulation objectives. The augmented timeline interface enabled in-depth context of trainee performances to be acquired, allowing coherent understanding of the trainee performance which was in sync with the visual observation. For feedback, the embedded video narration tool and drawing tool paired with the free camera for observation, enabled targeted and contextual feedback to be generated quickly, streamlining the assessment operation.

RQ7: *How useful is the framework compared to state-of-the-art methods for conducting authentic assessment and generating contextual feedback of room-scale virtual reality training simulations?*

RQ7 compared the themes from the framework, analytical data, and video assessment factors, examining their usefulness, usability and problems. Results demonstrated that analytical data was unsatisfactory for authentic assessment, due to its lack of reliability and validity in predefined learning outcomes, identified as the weakest factor throughout the evaluation. Consequently, analytical data hindered the involvement of the participant, diluting their authentic assessment and contextual feedback potential. The video assessment and feedback factor were both found to be viable, but, the framework demonstrated better usefulness due to the embedded authentic assessment and contextual feedback tools which streamlined the assessment and feedback process:

> *"I would go with the [the framework as first choice], because it can include the [features from] video assessment data. I think it does allow that nuance, there's no way I would pick [analytical data] on its own. [The framework] let's you see a little bit more. I saw the hesitancy in the applicant. And in a way [the framework] is cleaner [...] you've got their voice, but together with the visual aspect, you could see applicant was jumping backwards and forwards [... The framework] gave me a little bit more of an idea of their character." - (Participant 12)*

Although the video assessment demonstrated better usability due its familiarity (hosted on YouTube) and passive operation, participants noted many concerns which questions its viability long term, including potential for inducing motion-sickness, obstructed trainee performance data and losing track of trainee performance stages. Although the framework also experienced problems, most of the problems are resolvable, with most of the problems related to the presentation of the controls and technical interface. When comparing the framework to analytical data and video assessment, the framework improves upon the authentic assessment and contextual feedback process in almost all regards, given assessors have sufficient time to learn and become familiar with the controls and interface.

8.6.1 Four Key Findings

Four key themes emerged from the evaluation that are important for conducting authentic assessment in VR:

1. The inclusion of visual and audio cues are necessary during observation for assessors to acquire confidence and trust of trainees competency, along with validating their performance for tackling simulation objectives.
2. Assisted assessment and embedded feedback tools and features streamline the process and aid the assessor, as suggested by Hanoun and Nahavandi (2018).

3. Assessor observation from the trainee perspective improves the context acquired during assessment, allowing the assessor to better understand the logic and context of trainee actions. However, the perspective is unsuitable as the primary method of observation or assessment, as the erratic movement of the observation causes discomfort and motion sickness.

4. Follow-up dialogue sessions provide opportunities for the trainee and the assessors to check if the trainee has digested the feedback and learned from the experience, while also providing the opportunity for trainees to seek support from assessors.

In conclusion, the results from this evaluation suggest that the framework presented in this work is an effective, usable and useful method of providing assessment and generating feedback for assessors. Limitations of the framework documented concerns for the technical interface and control difficulties, caused by lack of familiarisation with the framework prior to the evaluation and network lag. However, even with these limitations, 11 out of 13 participants, ranked the framework as their first-choice for conducting authentic assessment in VR, showing assessors preferred the framework to the analytical data and video assessment factors. Chapter 9 summaries the results and contribution to knowledge from this work, concluding with potential directions for future work.

Chapter 9

Conclusion and Future Work

This chapter provides a summary of the conclusions of this PhD thesis based on the work preformed within the scope of the research project and research questions discussed in the previous chapters. This chapter presents the results of research questions (see Section 9.1), followed by the contributions to knowledge from this work (see Section 9.2), and concludes considerations for future work (see Section 9.3).

9.1 Conclusions

The objective of this research project was to design, develop and evaluate a framework for assessors, which would enable them to conduct authentic assessment and generate contextual feedback for room-scale VR training simulations. The framework was evaluated within a relevant training simulation focused on oil and gas training. However, due to the design of the framework, the components, tools and features are considered relevant for other fields of training and education where authentic assessment is routine, such as, but not limited to medical, aviation, and engineering. This is due to the scale-ability of the framework to capture all data from any simulations (that uses the Unity game engine) using the dual-recording methodology, which provides a consistent approach for assessor observation, assessment and generation of feedback, regardless of the simulation aims or objectives the framework is embedded into.

In this thesis, seven research questions were proposed to formulate the objective into phases of questions which outlines the research foundations this work was developed on.

These research questions were discussed in Chapter 3 and were formulated to determine:

RQ1 What is the state of the art for authentic assessment and contextual feedback of room-scale virtual reality training simulations?

RQ2 What evidence exists on usability and usefulness of identified approaches?

RQ3 What do assessors need for assessing and generating feedback for room-scale virtual reality training simulations?

RQ4 What tools are suitable to assess trainee performances in room-scale virtual reality training simulations?

RQ5 How can findings be compiled to create a framework for a generalised system for observation, assessment and feedback of room-scale virtual reality training simulations?

RQ6 Are the tools and features provided in the framework suitable for conducting authentic assessment and generating contextual feedback of room-scale virtual reality training simulations?

RQ7 How useful is the framework compared to state-of-the-art methods for conducting authentic assessment and generating contextual feedback of room-scale virtual reality training simulations?

The following sections will discuss each of the seven research questions, reviewing their main findings, contribution to knowledge and stressing the strengths and weaknesses in the research approach.

9.1.1 Methods of Assessment and Feedback of Virtual Reality Simulations

The first step of the research was to identify what existing methods are used to conduct authentic assessment or generate contextual feedback in virtual reality simulations and in the broader context of e-learning and virtual education. Because the research questions overlap in background research topics, two research questions were developed.

The first research question sought to discover *"What is the state of the art for authentic assessment and contextual feedback of virtual reality simulations?"* (RQ1). This research question identified the state-of-the-art for VR authentic assessment and determine whether a framework exists that assessors can use for authentic assessment and providing contextual feedback for trainees in room-scale VR.

Search terms for the review of the literature were determined to identify a wide but relevant selection of papers within the topics of virtual reality, game based learning and e-learning, focusing on the topics of remote assessment, feedback, training and observation. Due to still inconsistent naming conventions used by researchers in the field of VR for education, many papers discovered were not relevant, for example, virtual reality papers encapsulated a wide variety of technologies ranging from from desktop based, to room-scale virtual experiences, and everything in-between. Papers presented within the narrative structure of the literature were chosen due to their prominence in existing research and immediate relevance to the research questions of this work for room-scale virtual reality.

A narrative literature review was presented to produce a structured state-of-the-art overview covering a wide range of topics relating to the authentic assessment and contextual feedback of virtual training scenarios, including background research into the methodologies for conducting authentic assessment and contextual feedback. The review of existing literature did not reveal any existing framework suitable for conducting authentic assessment and generating contextual feedback in room-scale virtual reality simulations. Consequently, focus was directed on identifying existing methods of assessment and feedback that were used for virtual reality, e-learning, and remote education. Analysis of these methods resulted in a set of guidelines, see Section 2.9.

The second research question asked *"What evidence exists on usability and usefulness of identified approaches?"* (RQ2), and resulted in a review that discussed usability and usefulness of existing approaches for authentic assessment and contextual feedback in virtual reality room-scale simulations. The findings from the literature identified the strengths and weaknesses of existing approaches of assessment and feedback in virtual reality and relevant fields, including e-learning, remote education and distant learning. This research question was carried out to determine the design and features of the framework developed for this work. By investigating how existing methods were used for assessment and feedback in education and training, evidence was analysed to determine how the core useful and usable elements of these methods could be implemented into a single framework approach targeting room-scale virtual reality simulations.

Additional background research was carried out to provide the narrative structure to the literature review which present a coherent narration of the background problem and gap in the research, detailing the process of authentic assessment in virtual reality, the types of virtual training simulations, and discussing the importance of authentic assessment and contextual feedback in the learning process for the trainee.

However, as stressed by others (Chaudy, 2015), it cannot be guaranteed that every and all relevant papers were identified during the literature review. Given the wide variety of naming conventions for virtual reality and training, it is possible that some papers were not discovered or referenced within the searched databases.

9.1.2 Study with Industry Training Assessors

Following the findings from the literature review, which discovered that no framework existed for assessors to conduct authentic assessment and generate contextual feedback of room-scale virtual reality training simulations, a study was conducted with industrial training assessors to identify their requirements for a framework. The study was conducted to identify how classroom-based assessment is conducted, and determined how the identified authentic assessment and contextual feedback approaches could be transposed and improved within a virtual context. This identified the functional requirements of the framework (RQ3). Results from the study influenced the design of the observation, authentic assessment and contextual feedback components of the framework.

Results from RQ3 were used to generate the tools and features that best replicate the approaches identified by the study and meet assessors expectations conducting authentic assessment and generating contextual feedback (RQ4).

The study identified that authentic assessment needs direct observation of the trainees, without causing interference with the trainee performance or causing unnecessary observation anxiety (Shute, Ke, and Wang, 2017; Nyström et al., 2014) as is currently witnessed from classroom-based approaches (Nyström et al., 2014). The study also highlighted the need for tools and features to assist the assessor, allowing them to easily identify interesting or abnormal actions conducted by the trainee during their performance. Assisted assessment tools and features included summative performance indicators (see Section 6.5.1), an augmented timeline (see Section 6.5.2) and assisted visualisation of the trainee tools (see Section 6.5.3). Assessors who participated in the study discussed the requirements for contextual feedback, noting the importance of providing contextual and meaningful feedback that enables trainees to learn from.

The requirements identified from the study influenced the design and implementation of authentic assessment and contextual feedback tools that feature in the framework design, presented in Chapters 5 and 6, and which were implemented in Chapter 7.

Participant concerns of virtual training were reviewed during the study, revealing the perspectives from assessors on virtual training compared to traditional classroom-based approaches. Assessors are hesitant in adopting new technology, concerned about the potential of abuse and cheating of trainee performance outcomes, and about automated feedback, viewing it as unreliable for evaluating a trainee performance. Solutions to these concerns were incorporated into the design of the framework and components. However, some concern solutions need further work and research, Section 9.3.

In conclusion, the results from the study (see Section 4.2) influenced the design and development of the framework components, tools, and features.

9.1.3 Framework for Assessment and Feedback of Room-scale Virtual Trainee Simulations

After establishing existing methods for conducting authentic assessment and generating contextual feedback for virtual training simulations were inadequate for the requirements identified with industrial training assessors (see Section 4.2) and no existing framework was identified within the literature, the next question was to determine how a framework could be designed and developed (RQ5). The framework was designed to provide a seamless experience for assessors, allowing them to conduct authentic assessment and generate contextual feedback for trainee performances within a single framework environment. RQ5 asked *"How can findings be compiled to create a framework for a generalised system for observation, assessment and feedback of room-scale virtual reality training simulations?"* (RQ5), with the design of this framework and its components, tools, and features forming the main body of the contribution to knowledge from this work.

Because no existing framework was identified in the literature, the framework incorporated many elements from existing methods of conducting assessment and feedback, these approaches where examined against Hanoun and Nahavandi (2018) guidelines, to determine the overall design of a framework and the features required. These designs were then influenced and altered based on the findings from the study with industrial training assessors (see Section 4.1). Functional requirements from the study (see Section 4.2) were used to refine the functionality and tools embedded within the components of the framework, detailing the scope for how the framework operates.

The proposed framework consisted of seven components, with only five being the focus of this work. The framework incorporated an agile approach (Dingsøyr et al., 2012), enabling design and development stages to be modified and customised as components were created to form the foundation of the framework structure. The framework structure was designed to capture data, present it for direct observation and allow an assessor to conduct authentic assessment and generate contextual feedback for the trainee. Following the principles for generating feedback discussed in Section 1.6, the framework settled on the following components; Lightweight Recording, Processed Recording, Observation, Assessment and Feedback, with recording components using a novel dual-recording methodology due to the technical requirements necessary for the framework to meet the identified requirements from the study conducted with industrial training assessors (see Section 4.2).

The strength of this approach was the establishment of a framework that incorporated the findings from the background literature review of existing authentic assessment and contextual feedback approaches, which were incorporated with requirements identified from industrial training assessors to determine the design of the components and their functionality. However, this approach lacked input from the trainee who operates the training simulation and receives the feedback, this will be the focus of further work.

9.1.4 Usability and Usefulness Evaluation of the Framework

To conduct an evaluation of the the framework designed in Chapters 5 and 6, an implementation of the framework was created and discussed in Chapter 7 using the Unity Game Engine. The developed framework was then incorporated into a relevant commercial training simulation.

The first of the two remaining research questions asked *"Are the tools and features provided in the framework suitable for conducting authentic assessment and generating contextual feedback of room-scale virtual reality training simulations?"* (RQ6). This question was answered by evaluating the usefulness and usability of the framework to conduct authentic assessment and generated contextual feedback of a trainee performance during the evaluation. The second question *"How useful is the framework compared to state-of-the-art methods for conducting authentic assessment and generating contextual feedback of room-scale virtual reality training simulations?"* (RQ7) examined and compared the usefulness and usability of the framework against two existing methods of conducting assessment, video assessment and analytical data.

Because framework presented in this work is intended for assessors, and focuses on education and training. Participants were recruited from industry training and academic assessors, 6 participants from industry and 7 from academia. To evaluate to the framework, participants observed, assessed and generated feedback for two room-scale VR trainee performances, followed by a semi-structured interview and reflection on their experience of using the three methods.

The framework evaluation measured the usefulness and usability of the framework for conducting authentic assessment and generating contextual feedback of the trainee performance (RQ6), identifying the strengths and weaknesses of the framework. Findings from the evaluation (see Section 8.4) discovered that the framework tools and features for assisted assessment improved the ability for assessors to identify important elements of the trainee performance. For assessment, the controllable format of observation enabled the assessor to reliably monitor every action conducted by the trainee. Participants found feedback tools effective and valuable, allowing the creation of high-quality feedback information emboldened in the context of the trainee VR performance. The feedback supplied with the framework was considered equivalent to that provided in traditional classroom-based environments, which included context and targeted feedback. However, the usability of the controls were a concern, requiring long learning times due to unfamiliarity and network lag. Interface problems were exacerbated by controls difficulty, with participants having no prior experience using keyboard and mouse controls, suggesting the framework requires a more accessible interface and control scheme.

The final research question evaluated the framework against existing methods (RQ7) to determine if the framework improved upon them for conducting authentic assessment and generating contextual feedback. All three methods, i.e the framework implementation, video assessment and analytical data were evaluated by the 13 participants from industry and academia, determining how useful and usable they were, identifying each methods strengths and weaknesses. From the results it can be concluded that the framework was the most useful method, offering assisted assessment features and embedded feedback tools that streamlined the process and made the task of authentic assessment and contextual feedback easier. However, due to the lack of familiarisation of the framework visual layout and control interface, participants were forced to learn and search for common interface elements (Todi et al., 2018). These situations had assessors anticipate where they expected elements of the interface for the framework to be located (Todi et al., 2018). Because the video assessment factor was hosted on YouTube, with which participants are familiar, they were able to adapt to their assessment faster, and with less difficulty, making the video assessment method the second preferred method of assessors. However, 5 of the 13 participants highlighted that observing the video from the perspective of the trainee induced discomfort or motion-sickness due to the erratic and uncontrolled movement of the trainee. The least preferred method was analytical data. Although analytical data was useful to gain insight of when the trainee had issues in their performance, participants viewed it as unreliable and lacking context since they were not able to directly observe the trainee apply their knowledge to the tasks and objectives, which they viewed as crucial to conducting authentic assessment.

Results from the evaluation highlighted the importance of four key themes for providing feedback in virtual trainee simulations; (a) The requirement of observation of visual and audio cues of the trainee performance; (b) The need for assisted authentic assessment and contextual feedback tools to streamline the process and to offer greater confidence and control of the observation and assessment experience; (c) Assessor observation from the trainee perspective improves the context acquired during assessment, allowing the assessor to better understand the logic and context of trainee actions. However, the perspective is unsuitable as the primary method of observation or assessment, as the erratic movement of the observation causes discomfort and motion sickness, and; (d) Trainees with major gaps in knowledge should receive follow-up feedback through a contextualised question and answer dialogue session, allowing the assessor and the trainee to discuss and review the training performance together using visual and audio aids. These conversations provide confidence the trainee understands the feedback, and enables further opportunities for learning by reviewing any other concerns identified in the trainee performance during assessment.

In conclusion, the evaluation demonstrated that a streamlined framework that incorporates trainee simulation data to provide assisted assessment for the assessor is the most effective and useful method of observation. With the ability to access assisted tools and features, such as the summative performance indicators (see Section 6.5.1) and augmented timeline (see Section 6.5.2) during an assessment of a trainee performance, the assessor is able to acquire rich and detailed information, including visual and audio cues from direct observation and with complete control over the perspective and time of observation. The inclusion of feedback tools into the framework sped-up the generation of meaningful and contextual feedback for the trainee performance, which was considered equivalent to classroom-based feedback.

By discovering the usefulness and usability of the framework and existing methods from assessors perspectives, a clear indication of how authentic assessment and the generation of contextual feedback can be conducted in room-scale VR training simulations is discovered, answering the final two research question of this thesis. However, the evaluation was conducted remotely due to the Covid19 pandemic and it was impossible to run a face-to-face evaluation, running the evaluation remotely hindered the collection of some data, such as assessor gaze and body language.

9.2 Original Contribution to Knowledge

This thesis contributes to knowledge in different areas. First, this work has presented a new framework that identified and structures the components, tools, and features required for assessors to conduct authentic assessment and generate contextual feedback of room-scale virtual reality training simulations. Secondly, a novel dual-recording methodology allows run-time data capture and reconstruction of simulations that do not suffer from technical limitations that impact run-time performance (Howie and Gilardi, 2021).

The third contribution to knowledge is the evaluation conducted with assessors from industry and academia. The framework and existing methods, such as analytical data and video assessment, were used determine the strengths and weaknesses related to their usability and usefulness for conducting authentic assessment and generating contextual feedback in room-scale VR simulations.

The literature review in Chapter 2 presented several gaps that this thesis contributes to close: (a) lack of a framework design for assessors to conduct authentic assessment and generate contextual feedback of room-scale virtual reality training performances; (b) existing embedded recording techniques for VR simulations have negative impact on run-time performance, or fail to capture sufficient data for assessment to be conducted; (c) lack of information for what assessors require for authentic assessment and contextual feedback for virtual trainee performances; (d) how would a streamlined framework compare against existing methods for conducting authentic assessment and generating contextual feedback in room-scale virtual reality training simulations.

The objective of this research was to rectify the gap in literature, by designing, developing and evaluating a framework that enables assessors to conduct assessment and generate contextual feedback. Existing methods of conducting assessment were unsuitable for the requirements of room-scale virtual reality, since the conditions for capturing data could not be guaranteed and did not meet the functional requirements identified from the study with industrial training assessors (see Section 4.2).

This work is grounded in the literature building upon existing elements of work from Raij and Lok (2008), Stone, Snell, and Cooke (2016), (Backlund et al., 2007) and Thurston and Martin (2011). These applications are not similar to the framework presented in this work, but demonstrate ideas and implementations which influenced the design of components for conducting authentic assessment and generating contextual feedback. Furthermore, the framework followed the guidelines of after-action review by Hanoun and Nahavandi (2018).

The originality of this work stems from the novelty of the framework and presentation of the requirements necessary for assessors to perform authentic assessment in room-scale virtual reality, both of which have been identified as a gap within knowledge (Hamilton et al., 2021; Adams et al., 2021).

9.2.1 Publications

The author has published and presented the following papers in peer-reviewed conferences and journals:

Howie, S.R. and Gilardi, M., 2019, May. Virtual Observation of Virtual Reality Simulations. In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (pp.

1-6).

Howie, S., Gilardi, M. Virtual Observations: a software tool for contextual observation and assessment of user's actions in virtual reality. Virtual Reality (2020). https://doi.org/10.1007/s10055-020-00463-5

Howie, S., Gilardi, M. Methods and Systems for Recording a User Experience. UKIPO Patent (Filed) GB2104444.1, March. 2021

Howie, S., Gilardi, M. Design and Development of Authentic Assessment Tools for Room-scale Virtual Reality Training. Virtual Reality (Submitted, expected 2021).

9.2.2 Requirements for Developing an Authentic Assessment and Contextual Feedback Framework for Virtual Reality Simulations

Assessors were the targeted audience of the framework, allowing them to directly observe room-scale VR simulation performances, conduct authentic assessment and generate contextual feedback using a streamlined all-in-one framework.

Chapter 4 of this work presented a study with industrial training assessors which formed the requirements of components, tools, and features within the framework. The study was the foundation of the design of the five components that are the focus of this work; Lightweight Recording, Processed Recording, Observation, Assessment and Feedback. Chapters 5 and 6 present the design, with Chapter 7 presenting the technical implementation of the framework, detailing the novel dual-recording methodology, the structural operation of the components and tools embedded into the Unity Game Engine.

The framework was developed and evaluated in a relevant training environment. By evaluating the framework against existing methods for conducting assessment, the findings contribute to the academic conversation on the subject that researches can examine or build upon. Additionally, the evaluation demonstrated the importance of the three four components of a framework concluding; visual and audio cues are crucial for the assessment, streamlined authentic assessment and contextual feedback tools improve the usefulness and usability of the framework, observation from the trainee perspective improves the context acquired during assessment, but its erratic movement induces discomfort and motion-sickness in assessors, and follow-up contextual dialogue sessions should be conducted for trainees who demonstrate severe lack of knowledge in training operations.

9.2.3 Dual Recording Methodology

Existing recording techniques were unable to capture all simulation data without impacting the run-time performance of the trainee. This thesis contributes to the knowledge for the process of capturing all simulation trainee data, using a dual-recording methodology, splitting the recording to two components; Lightweight Recording and Processed Recording. By capturing all simulation data, deep insight into the trainee performance can be examined and used to provide assisted assessment tools and features which aid the assessor and improve the process of conducting assessment.

This work contributes to knowledge by presenting the design and implementation of a dual-recording methodology, using a lightweight recording component to capture minimal trainee data during run-time, followed by a secondary processed recording component which captures all remaining data. This methodology is universally applicable to all recording operations by splitting variable data into three categories: User Input, Deterministic and Non-deterministic. The data output by the recording components is a complete capture of a trainee performance, including all variable data values throughout the training experience. These complete data-captures provide many interesting directions of future research, most interestingly, the application of automatic assessment from artificial intelligence which could evaluate the data and present additional insight into the experience.

9.2.4 Embedded Framework Tools for Assessment and Feedback in Virtual Reality

By embedding all necessary tools for authentic assessment and contextual feedback into the framework, the assessment and feedback of the trainee is streamlined, allowing the assessor to review trainee performances in-depth, highlighting the shortcomings in the trainee performance.

Moreover, the assisted assessment features allows assessors to quickly identify areas in the simulation that were important to determining the outcome of the trainee performance. Findings from the evaluation suggest that the streamlining of the authentic assessment and contextual feedback into an all-in-one framework enables an effective approach for conducting authentic assessment and generating contextual feedback. This approach avoids cognitive overload for the assessor, as their attention and focus is fixed, rather than being diverted between different applications, as was the case when using analytical data and video assessment methods.

The framework evaluation demonstrated the potential for assisted assessment features using summative performance indicators and an augmented timeline interface to assist assessors. Using the data available from the recording components, the assisted assessment features can be designed to suit the needs of the assessor. For example, Stone, Snell, and Cooke (2016) demonstrated an assisted feature in their after-action review implementation, overlaying a divers search path during a training simulation to aid the assessment experience. However, Stone, Snell, and Cooke, 2016 after-action review was designed to capture the search path data during the run-time simulation. With the framework, all of the simulation data is captured and reproduced, making it possible for these features to be included and built-upon overtime. It is envisioned that with future work, these features for creating assisted assessment features will be aided by the incorporation of the artificial intelligence, further expanding the embedded and assisted assessment features.

The evaluation of these tools embedded into the framework provide other researchers and developers with the knowledge of component requirements and their usefulness. This allows the replication of tools and functionalities that enable assessors to have confidence in their ability to directly observe the trainee, conduct assessment and provide feedback

In summary, the framework demonstrated in this thesis forms the components required for assessors to conduct authentic assessment and generate contextual feedback for room-scale VR training simulations. Other components for live streaming and artificial intelligence will be considered in future work, with the components expanding the potential of the framework.

9.3 Future Research Directions

During the development of the framework and its components, tools, and features, and their evaluation, many areas for improvement and future research were identified. This section will detail six of these areas: improvement of the user interface and control scheme to decrease complexity; the transcription of verbal audio captured from the trainees microphone during their performance; how an additional automatic assessment system could be incorporated into the framework to assist assessors; incorporate a live component for multi-user or collaborative assessment; incorporation of eye tracking to add additional context of the trainee gaze; and a study to determine effectiveness and usability of the contextual feedback generated from the framework with trainees.

Improvements for the Framework

The biggest concern from the evaluation of the framework was the usability issues caused by the unfamiliar user interface and difficult control scheme for operating the free camera tool. These two concerns are immediate future research directions, required to determine how the interface and controls can best present the data to assessors in a manner that makes the direct observation, assessment and generating feedback easier to achieve.

For the user interface, work by Todi et al. (2018) demonstrates potential avenues of improvement for the framework, generating the layout of the interface to match the end-users history of interfaces. While the interface of the framework designed and developed in Chapters 6 and 7 was designed to replicate the overall structure of video editing applications, non-technical users were unable to acclimatise to the interface quickly and struggle to comprehend all the features and functionality. These concerns would likely improve as end-users are given more time to train and operate the framework interface, but this solutions do not neglect the fact that the framework in its current state presents accessibility problems. Future research should be conducted to determine how the free camera tool controls are best suited to enabling assessors the ability to interact and engage with the observation, authentic assessment and contextual feedback components of the framework from a usability perspective.

However, usability problems were not restricted to the user interface. The control scheme for operating the free camera tool was a main source of frustration for assessors during their evaluation of the framework. The frustration was caused by a variety of methods surrounding the implementation of the free camera tool controls, network lag, and the challenges of participants adopting and connecting the visuo-motor control of the camera with keyboard and mouse controls (Green and Bavelier, 2006). Future work should focus on alternative implementations of the camera control and perspective functionality. One suggestion by participants during the study was to have predefined camera perspectives, allowing the assessors to focus their attention on the trainee and detaching them from the task of moving the free camera tool. In this example, predefined camera positions would be generated at the action or objective areas for the simulation, allowing the assessor to change perspective easily. However, participants did note that they would still prefer to have the ability to have full control of the camera perspective, but utilise it only as necessary and as a secondary advanced feature. It is anticipated that future work regarding a study of different implementations of camera controls would be of interest for the wider research community, providing further requirements for the requirements of observation and assessment in virtual simulations.

Beyond future research directions for user interface and control schemes, there are general framework improvements that were discussed or hinted at by participants during their reflection of the evaluation. These are more focused on the framework implementation created in Chapter 7, but shed light on the direction of other implementations of the framework presented in this work. These general framework improvements include; additional annotation ability beyond manual drawing and annotating, including transparent highlighting and other features found in common photo-editing suites; embedding a preview view for the augmented timeline modification to replicate YouTube's small preview video on the augmented timeline, enabling assessors to quickly visually identify key-points within the simulation, and; the incorporation of written records into the framework, used as both a method of tagging key frames in recorded data, as done by Thurston and Martin's (2011) GEAARS interface for future observations, and to enable written comments and logs to be attached to the feedback of a trainee performance, allowing for integration with existing trainee record and archival systems.

It is also envisioned that in future work, the end-user of the framework conducting the assessment would be able to query data to create dynamic findings from analysis of the variable data. For example, the end-user assessor could query the framework to identify if and where the trainee variable information matched a set of inputted conditions, taking inspiration from non-developer friendly programming block logic (Howland and Good, 2015). However, for the framework implementation, the summative information of variable data was designated by default to display the true or fail outcomes for the objectives listed in Figure 6.7, however, any information from the captured data could have been queried and presented, the simulation objectives were chosen as they were considered most relevant for the simulation scenario used in the evaluation.

Ideally, the augmented timeline (see Section 6.5.2) would be dynamically modified to change the variables displayed, with an assessor being able to search for variable names or click on objects in the 3D environment and select variables to monitor, providing better flexibility of the observation. However, due to time restrictions, these features were not implemented. It is envisioned that in future work, the ability to access variable information of all simulation can lead to interesting expansions with regards to assessment aids.

Finally, the room-scale tracking feature should be evaluated with assessors to determining its usefulness for acquiring information of the trainee performance. It is envisioned that the feature is an important element of conducting assessment due to the prominence of problems with room-scale VR simulations with regards to limited physical space in the real-world impacting the capabilities of VR actions and movement.

Transcription and Presentation of Trainee Audio

One of the core findings of this work was demonstrating how important both visual and audio cues were on the estimation of confidence level of the trainee. While the visual concerns are mostly addressed in section 9.3, discussing the inclusion of a visual timeline preview, however, audio cues are not immediately noticeable and require a different approach. An inspiration for future work is the work by Raij and Lok (2008) who transcribed participant audio and embedded the timestamped markings to points within the simulation, allowing timestamps to be selected and observed from that point of the recording. Future work should consider functionality to auto-transcribe the trainee data, and investigate how to best present the audio verbal cues of the trainee in sync with the other data presented on the interface, including the data augmented timeline and summative information. This future work should improve the viability of skipping sections of trainee performance, using the audio transcription of the trainee and recording of variable information to determine areas of the trainee performance which has no assessment or observation value. This direction of future work should determine the technical capability of transcribing the data, but also further investigate how the data offered by the transcription could be used for assessment, both by assessors, and its incorporation into the artificial intelligence component.

Automatic Assessment System

The findings from literature demonstrated why automatic assessment is not yet capable of replacing assessors. One of the findings of this research is that assessors would not trust automatic outcomes without having witnessed the trainee directly apply their logical knowledge to the objectives. However, this does not necessarily mean that automatic assessment systems are not viable, instead, an automatic assessment system should (at-present) be utilised for two methods: (a) To provide an estimation of the trainee performance immediately to the trainee, giving an immediate estimated indication of how their performance outcomes were measured against other trainees. While the automatic assessment may not be able to collect and correctly analyse all the nuances of the trainee performance, such as audio cues, the estimation should be able to determine an overall category of the training performance for pass or fail. The estimation can then be validated and modified based on the direct observation of an assessor, who would conduct the in-depth assessment and generate the necessary feedback. (b) Secondly, automatic assessment should be used to assist the assessor during the assessment of the trainee, allowing interesting areas of the trainee performance to be pre-identified, improving and streamlining the process of assessor observation of the trainee. This feature would tag, log and mask areas of the trainee performance using observation and assessment that are known to be incorrect or invalid actions, taking inspiration from the automated systems of Action-Based assessment discussed in Section 2.7.5.

The reasons for conducting future work into providing immediate estimated feedback to a trainee performance is inspired from discussions with assessors. During the initial study in Section 4.1 and the evaluation in Chapter 8, assessors spoke of the importance to minimising the gap between conducting assessment and receiving feedback. The literature back these findings (Wijewickrema et al., 2017), stressing that feedback should be generated and sent to the trainee as soon as possible. However, because one of the main advantages of virtual reality is the ability to perform the simulations any time at any location, it is unlikely that feedback from an assessor will be instantaneous. Future work should investigate how the data captured during the recording components can be used to dynamical evaluate the trainee performance to provide an immediate estimation of assessment outcomes to the trainee. These estimations can be later validated and expanded by the assessors, who can directly observe the trainee to determine if the success of the simulation was due to luck or if the trainee demonstrated a logical application of knowledge.

When envisioning an automatic assessment system to assist the assessor, it is in the author's opinion that future work on the system should expand upon the ideas of augmented timeline functionality demonstrated in this work. By embedding artificial intelligence into the recognition of patterns from the modification of variable data, it may be possible to determine a recognition system to automatically identify unexpected actions, responses or verbal cues (transcribed trainee verbal audio cues) that indicated important or interesting sections of the simulation. Furthermore, since all data from the training simulation is captured by the dual-recording methodology, development of assisted assessment tools and features are not restricted to pre-training simulations. As a consequence, automatic assessment aids can gradually be improved iteratively as future work projects. Eventually, future work in automatic assessment may be able to replace the role of the assessor, however, it is important to stress that even with an automatic assessment system conducting the authentic assessment and contextual feedback, the trainee performance should be recorded and archived to enable validation by an assessor if necessary. Otherwise, a flaw or bug for the automatic assessment system could lead to fatal consequences of trainees being incorrectly advised, leading to real-world disasters.

Live Multi-User Assessment and Feedback

Another avenue for future work is to expand upon the demonstration of live observation of trainee performances from (Howie and Gilardi, 2020). The publication demonstrates how multi-user observation of trainee performances can be observed live, with data encoded and saved to the all computer devices (Howie and Gilardi, 2020). Future work could incorporate authentic assessment and contextual feedback tools directly into the live-assessment environment, creating the possibility for immediate feedback to the trainee and providing the opportunity for live interactive dialogue sessions between the assessor and trainee(s). Depending on the directions of future research, this work could expand the role of the framework beyond after-action authentic assessment and contextual feedback, incorporating proactive teaching and education components in live interactive and discussion sessions that replicate classroom-based teaching. It is envisioned that future work in this direction would enable immediate authentic assessment and contextual feedback, with both the assessor and trainee(s) being present within a single simulation environment, potentially expanding the framework beyond authentic assessment and contextual feedback and into the realm of teaching and educating.

Eye-Tracking Gaze

The potential for assisting observation of the assessment with eye-tracking is discussed in (Howie and Gilardi, 2020). Within the components of the framework, eye-tracking makes it possible for the assessor to be able to identify what the trainee is looking at and determine how that relates to their actions during the simulation, similar to work by Thanyadit, Punpongsanon, and Pong (2020). This feature would be most suited to areas of a trainee performance where the trainee is struggling or failing objectives, providing deeper insight into a trainees movement and eye-gaze relative to their actions.

Eye-tracking was embedded into the assessment component of the framework to assist the assessor better understand the gaze of the trainee during their performance. However, the VR equipment necessary to capture eye-tracking data was inaccessible for the evaluation of the framework. In future work, it is envisioned that this feature would be expanded with the augmented timeline interface previously discussed in Section 6.5.2, allowing better clarification of gaze triggered objects, allowing the assessor to query the data captured from the simulation to look at instances in the simulation when the trainee was looking at specific areas or items during their training performance.

Research with Trainees

The final suggestion for future work is for further research to be conducted from the perspective of trainees. This work focused on the perspective of assessors who conduct the authentic assessment and contextual feedback, and as mentioned in the scope of the research (see Section 1.5), it was not possible to involve trainees in the research approach. Future work should be conducted with the trainees to determine their role in the operating the training simulation, and the usefulness and usability of the feedback generated by the assessors. There are various avenues for future work, including the content, context and accessibility of the feedback, and how the feedback provided by the framework compares to real-world feedback from classroom-based approaches. Future studies with trainees could gauge the process of the receiving feedback, and determine if they consider follow-up dialogue sessions as an important element, as was stressed by assessors in the findings of this work (see Section 8.3.4). However, the core objective of this future research direction is to involve the trainee in the framework process and identify how they can be incorporated into the feedback component of this framework.

9.4 Protecting User Privacy

On a final note for this thesis, it is important to stress the negative consequences of capturing and reproduction virtual reality user data with regards to protecting their privacy. The potential for abuse of a digital recording methodology is evident, with significant insight available into the trainees movement, eye-gaze, voice, actions, reactions and general interactivity, action must be taken to protect and secure all captured data from a trainee. Miller et al. (2020) captured participant data and then trained an automated system to identify unique characteristics of each individual. The system was able to identify each trainee with 95% accuracy with only five minutes of tracking data per person, demonstrating the richness of VR data. It is in the belief of this author (Howie and Gilardi, 2019) and of others (Miller et al., 2020; Sharma et al., 2017) that extreme care is taken to prevent the abuse of data made accessible from VR equipment. Notably, the use of the software methodology described in this work has demonstrated the ability to capture and reproduce substantial information from a user operating within VR equipment. It is important that users who participate in using this methodology are aware of the data being captured and stored, and consent to the data being used. It is imperative that the data captured should not be used for illicit purposes, or commercialisation gains without given consent by the data source (trainee or VR user). A final recommendation of this work is to follow that of Miller et al. (2020) and advise that the capture of VR data is made explicitly aware and consented to by all users due to be recorded. This information and consent should be ethically conducted, prior to any data being captured or stored. Where possible, all data should be kept anonymous to prevent any individual from being identified should data be leaked, hacked or exploited.

References

Abdelaal, Alaa Eldin et al. (2020). "A multi-camera, multi-view system for training and skill assessment for robot-assisted surgery". In: *INTERNATIONAL JOURNAL OF COMPUTER ASSISTED RADIOLOGY AND SURGERY*.

Abidi, Mustufa Haider et al. (2019). "Assessment of virtual reality-based manufacturing assembly training system". In: *The International Journal of Advanced Manufacturing Technology* 105.9, pp. 3743–3759.

Abraham, Reina M and Veena S Singaram (2019). "Using deliberate practice framework to assess the quality of feedback in undergraduate clinical skills training". In: *BMC medical education* 19.1, p. 105.

Abuhammad, Areej et al. (2021). ""MedChemVR": A Virtual Reality Game to Enhance Medicinal Chemistry Education". In: *Multimodal Technologies and Interaction* 5.3. ISSN: 2414-4088. DOI: 10.3390/mti5030010. URL: https://www.mdpi.com/2414-4088/5/3/10.

Adams, Andrea et al. (2021). "Potentials of Teaching, Learning, and Design with Virtual Reality: An Interdisciplinary Thematic Analysis". In: *Intersections Across Disciplines: Interdisciplinarity and learning*. Ed. by Brad Hokanson et al. Cham: Springer International Publishing, pp. 173–186. ISBN: 978-3-030-53875-0. DOI: 10.1007/978-3-030-53875-0_14. URL: https://doi.org/10.1007/978-3-030-53875-0_14.

Agreval, Sarvesh et al. (2019). "Defining immersion: Literature review and implications for research on immersive audiovisual experiences". In: *AES 147th Convention*. Audio Engineering Society, p. 10275.

Alruwais, Nuha, Gary Wills, and Mike Wald (2018). "Advantages and challenges of using e-assessment". In: *International Journal of Information and Education Technology* 8.1, pp. 34–37.

Andersen, Steven Arild Wuyts et al. (2019). "The effect of structured self-assessment in virtual reality simulation training of mastoidectomy". In: *European Archives of Oto-rhino-laryngology* 276.12, pp. 3345–3352.

Andreatta, Pamela B et al. (2010). "Virtual reality triage training provides a viable solution for disaster-preparedness". In: *Academic emergency medicine* 17.8, pp. 870–876.

Armas, Claudia de, Romero Tori, and Antonio Valerio Netto (2020). “Use of virtual reality simulators for training programs in the areas of security and defense: a systematic review”. In: *Multimedia Tools and Applications* 79.5, pp. 3495–3515.

Ashford-Rowe, Kevin, Janice Herrington, and Christine Brown (2014). “Establishing the critical elements that determine authentic assessment”. In: *Assessment & Evaluation in Higher Education* 39.2, pp. 205–222.

Backlund, Per et al. (2007). “Sidh–a game based firefighter training simulation”. In: *2007 11th International Conference Information Visualization (IV’07)*. IEEE, pp. 899–907.

Bayerl, P Saskia et al. (2019). “Evaluating Serious Game Trainings”. In: *Serious Games for Enhancing Law Enforcement Agencies*. Springer, pp. 149–169.

Beller, Michal (2013). “Technologies in Large-Scale Assessments: New Directions, Challenges, and Opportunities”. In: *The Role of International Large-Scale Assessments: Perspectives from Technology, Economy, and Educational Research*. Ed. by Matthias von Davier et al. Dordrecht: Springer Netherlands, pp. 25–45. ISBN: 978-94-007-4629-9. DOI: 10.1007/978-94-007-4629-9_3. URL: https://doi.org/10.1007/978-94-007-4629-9_3.

Bennett, Randy Elliot (2015). “The Changing Nature of Educational Assessment”. In: *Review of Research in Education* 39.1, pp. 370–407. DOI: 10.3102/0091732X14554179. eprint: https://doi.org/10.3102/0091732X14554179. URL: https://doi.org/10.3102/0091732X14554179.

Bernard, Fabien et al. (2018). “Virtual reality simulation and ergonomics assessment in aviation maintainability”. In: *Congress of the International Ergonomics Association*. Springer, pp. 141–154.

Berndt, Henrik et al. (2018). “Immersion and Presence in Virtual Reality Training for Mass Casualty Incidents.” In: *ISCRAM*.

Bertram, Johanna, Johannes Moskaliuk, and Ulrike Cress (2015). “Virtual training: Making reality work?” In: *Computers in Human Behavior* 43, pp. 284–292.

Billings, DR (2012). “Efficacy of adaptive feedback strategies in simulation-based training”. In: *Military Psychology* 24.2, pp. 114–133.

Black, Paul and Dylan Wiliam (1998). “Assessment and classroom learning”. In: *Assessment in Education: principles, policy & practice* 5.1, pp. 7–74.

Bliss, James P, Philip D Tidwell, and Michael A Guest (1997). “The effectiveness of virtual reality for administering spatial navigation training to firefighters”. In: *Presence: Teleoperators & Virtual Environments* 6.1, pp. 73–86.

Blomberg, Jeanette, Mark Burrell, and Greg Guest (2007). “The Human-computer Interaction Handbook”. In: ed. by Julie A. Jacko and Andrew Sears. Hillsdale, NJ, USA: L. Erlbaum Associates Inc. Chap. An Ethnographic Approach to Design, pp. 964–986. ISBN: 0-8058-3838-4. URL: http://dl.acm.org/citation.cfm?id=772072.772133.

Boud, David (1995). “Assessment and learning: contradictory or complementary”. In: *Assessment for learning in higher education*, pp. 35–48.

Bozkir, Efe, David Geisler, and Enkelejda Kasneci (2019). “Assessment of driver attention during a safety critical situation in VR to generate vr-based training”. In: *ACM Symposium on Applied Perception 2019*, pp. 1–5.

Braun, Virginia and Victoria Clarke (2006). “Using thematic analysis in psychology”. In: *Qualitative Research in Psychology* 3.2, pp. 77–101. DOI: 10.1191/1478088706qp063oa. eprint: https://www.tandfonline.com/doi/pdf/10.1191/1478088706qp063oa. URL: https://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa.

— (2019). “To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales”. In: *Qualitative Research in Sport, Exercise and Health* 0.0, pp. 1–16. DOI: 10.1080/2159676X.2019.1704846. eprint: https://doi.org/10.1080/2159676X.2019.1704846. URL: https://doi.org/10.1080/2159676X.2019.1704846.

Bredl, Klaus et al. (2015). “The Avatar as a Knowledge worker? How immersive 3D virtual environments may foster knowledge acquisi-tion”. In: *Leading Issues in Knowledge Management, Volume Two* 2, p. 222.

Bric, Justin D et al. (2016). “Current state of virtual reality simulation in robotic surgery training: a review”. In: *Surgical endoscopy* 30.6, pp. 2169–2178.

Brooke, John (1996). “SUS: a “quick and dirty’usability”. In: *Usability evaluation in industry*, p. 189.

Broquet, Karen and Mantosh Dewan (2016). “Evaluation and feedback”. In: *International Medical Graduate Physicians*. Springer, pp. 41–55.

Broquet, Karen E and Manisha Punwani (2012). “Helping international medical graduates engage in effective feedback”. In: *Academic Psychiatry* 36.4, pp. 282–287.

Brown, Emily and Paul Cairns (2004). “A grounded investigation of game immersion”. In: *CHI’04 extended abstracts on Human factors in computing systems*, pp. 1297–1300.

Bucarelli, Nidia, Jiansong Zhang, and Chao Wang (2018). “Maintainability assessment of light design using game simulation, virtual reality and brain sensing technologies”. In: *Proceedings of ASCE Construction Research Congress*, pp. 378–387.

Buttussi, Fabio and Luca Chittaro (2017). “Effects of different types of virtual reality display on presence and learning in a safety training scenario”. In: *IEEE transactions on visualization and computer graphics* 24.2, pp. 1063–1076.

Carruth, Daniel W (2017). “Virtual reality for education and workforce training”. In: *2017 15th International Conference on Emerging eLearning Technologies and Applications (ICETA)*. IEEE, pp. 1–6.

Charissis, Vassilis et al. (June 2008a). "3D stereoscopic design of submersive rescue vehicle and rescue mission simulation". English. In: *International Conference of Warship 2008: Naval Submarines 9*. Abstract only available from publisher's website js 16.05.13. The Royal Institution of Naval Architects, pp. 13–20. ISBN: 9781905040476.

Charissis, Vassilis et al. (2008b). "An enquiry into VR interface design for medical training: VR augmented anatomy tutorials for breast cancer". In: *The Engineering Reality of Virtual Reality 2008*. Vol. 6804. International Society for Optics and Photonics, p. 680404.

Chaudy, Yaelle (Dec. 2015). "An Assessment and Learning Analytics Engine for Games-based Learning". PhD thesis. DOI: 10.13140/RG.2.1.4932.5040.

Chavez, Bayron and Sussy Bayona (2018). "Virtual Reality in the Learning Process". In: *Trends and Advances in Information Systems and Technologies*. Ed. by Álvaro Rocha et al. Cham: Springer International Publishing, pp. 1345–1356. ISBN: 978-3-319-77712-2.

Chen, Chwen Jen, Seong Chong Toh, and Wan Mohd Fauzy Wan Ismail (2005). "Are learning styles relevant to virtual reality?" In: *Journal of research on technology in education* 38.2, pp. 123–141.

Chilimbi, Trishul et al. (2014). "Project adam: Building an efficient and scalable deep learning training system". In: *11th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 14)*, pp. 571–582.

Clark, Dan and Jeff Sanders (2011). *Beginning C# object-oriented programming*. Springer.

Collins, Ayşe Baş (2002). "Gateway to the real world, industrial training: Dilemmas and problems". In: *Tourism management* 23.1, pp. 93–96.

Collins, Hilton (2018). *Video Game Puts the 'Edge' in First Responder Training*. URL: http://www.govtech.com/videos/Video-Game-Edge-First-Responder-Training.html.

Cooper, Natalia et al. (2018). "The effects of substitute multisensory feedback on task performance and the sense of presence in a virtual reality environment". In: *PloS one* 13.2, e0191846.

Cordeil, Maxime et al. (2017). "Design space for spatio-data coordination: Tangible interaction devices for immersive information visuali-sation". In: *Pacific Visualization Symposium (PacificVis), 2017 IEEE*. IEEE, pp. 46–50.

Danforth, Douglas R et al. (2009). "Development of virtual patient simulations for medical education". In: *Journal For Virtual Worlds Research* 2.2.

Davis, Walter D (2005). "The interactive effects of goal orientation and feedback specificity on task performance". In: *Human Performance* 18.4, pp. 409–426.

Deeley, Susan J (2018). "Using technology to facilitate effective assessment for learning and feedback in higher education". In: *Assessment & Evaluation in Higher Education* 43.3, pp. 439–448.

Dingsøyr, Torgeir et al. (2012). *A decade of agile methodologies: Towards explaining agile software development*.

Dunne, James R and Claudia L McDonald (2010). "Pulse!!: a model for research and development of virtual-reality learning in military medical education and training". In: *Military medicine* 175.suppl_7, pp. 25–27.

El Boghdady, Michael and Afshin Alijani (2017). "Feedback in surgical education". In: *The Surgeon* 15.2, pp. 98–103.

Ellaway, Rachel and Ken Masters (2008). "AMEE Guide 32: e-Learning in medical education Part 1: Learning, teaching and assessment". In: *Medical teacher* 30.5, pp. 455–473.

Epic Games (2021a). *Replay System*. (Accessed on 16/01/2021). URL: `https://docs.unrealengine.com/en-US/Engine/Replay/Streamers/index.html`.

— (2021b). *The most powerful real-time 3D creation platform*. (Accessed on 16/01/2021). URL: `https://www.unrealengine.com/en-US/`.

Eubanks, James Coleman et al. (2016). "Full-body portable virtual reality for personal protective equipment training". In: *International Conference on Virtual, Augmented and Mixed Reality*. Springer, pp. 490–501.

EveReal (2021). *Video Capture Pro*. (Accessed on 16/04/2021). URL: `https://assetstore.unity.com/packages/tools/video/video-capture-pro-155663`.

Fardinpour, Ali and Heinz Dreher (2012). "Towards developing automated assessment in virtual learning environments". In: *IADIS International Conference on Internet Technologies & Society*.

Fardinpour, Ali, Torsten Reiners, and Lincoln C Wood (2018). "Action-based learning assessment in virtual training environments". In: *Authentic Virtual World Education*. Springer, pp. 79–94.

Farra, Sharon et al. (2019). "Effects of virtual reality simulation on worker emergency evacuation of neonates". In: *Disaster medicine and public health preparedness* 13.2, pp. 301–308.

Farra, Sherry et al. (2016). "Storyboard Development for Virtual Reality Simulation". In: *Clinical Simulation in Nursing* 12.9, pp. 392–399.

FEMA (Mar. 2018). *FEMA National Excerise Simulation Center*. "`https://www.movesinstitute.org/wp-content/uploads/2012/07/FEMA-National-Excerise-Simulation-Center-LegaryMahrous1.pdf`. (Accessed on 06/03/2018).

FitzGerald, Elizabeth (2012). "Analysing video and audio data: Existing approaches and new innovations". In: *Open Reserach Online*.

Fossey, Ellie et al. (2002). "Understanding and evaluating qualitative research". In: *Australian & New Zealand Journal of Psychiatry* 36.6, pp. 717–732.

Fukkink, Ruben Georges, Noortje Trienekens, and Lisa JC Kramer (2011). "Video feedback in education and training: Putting learning in the picture". In: *Educational Psychology Review* 23.1, pp. 45–63.

Ganni, Sandeep et al. (2018). "A software-based tool for video motion tracking in the surgical skills assessment landscape". In: *Surgical endoscopy* 32.6, pp. 2994–2999.

Gielen, Sarah, Filip Dochy, and Sabine Dierick (2003). "Evaluating the consequential validity of new modes of assessment: The influence of assessment on learning, including pre-, post-, and true assessment effects". In: *Optimising new modes of assessment: In search of qualities and standards*. Springer, pp. 37–54.

Goderstad, JM et al. (2019). "Development and validation of a general and easy assessable scoring system for laparoscopic skills using a virtual reality simulator". In: *European Journal of Obstetrics & Gynecology and Reproductive Biology: X* 4, p. 100092.

Goldberg, Stephen L, Bruce W Knerr, and James Grosse (2003). *Training dismounted combatants in virtual environments*. Tech. rep. Army Research Development and Engineering Command Orlando FL Simulation ...

Gould, Dinah et al. (2001). "Examining the validity of pressure ulcer risk assessment scales: developing and using illustrated patient simulations to collect the data INFORMATION POINT: Visual Analogue Scale". In: *Journal of clinical nursing* 10.5, pp. 697–706.

Green, C Shawn and Daphne Bavelier (2006). "The cognitive neuroscience of video games". In: *Digital media: Transformations in human communication* 1.1, pp. 211–223.

Greenhalgh, Chris, Jim Purbrick, and Dave Snowdon (2000). "Inside MASSIVE-3: flexible support for data consistency and world structuring". In: *Collaborative Virtual Environments: Proceedings of the third international conference on Collaborative virtual environments*. Vol. 2000. Citeseer, pp. 119–127.

Greenhalgh, Chris et al. (2002). "Applications of temporal links: Recording and replaying virtual environments". In: *Proceedings IEEE Virtual Reality 2002*. IEEE, pp. 101–108.

Grübel, Jascha et al. (2017). "EVE: A Framework for Experiments in Virtual Environments". In: *Spatial Cognition X*. Ed. by Thomas Barkowsky et al. Cham: Springer International Publishing, pp. 159–176. ISBN: 978-3-319-68189-4.

Gulikers, Judith, Theo Bastiaens, and Paul Kirschner (2006). "Authentic assessment, student and teacher perceptions: the practical value of the five-dimensional framework". In: *Journal of Vocational Education and Training* 58.3, pp. 337–357.

Gulikers, Judith TM, Theo J Bastiaens, and Paul A Kirschner (2004). "A five-dimensional framework for authentic assessment". In: *Educational technology research and development* 52.3, p. 67.

Gunn, Therese et al. (2018). "The use of virtual reality simulation to improve technical skill in the undergraduate medical imaging student". In: *Interactive Learning Environments* 26.5, pp. 613–620.

Guterman, Pearl S (2005). "Psychological preparedness for disaster". In: *Academia. edu.[online] Available at:* `http://yorku.academia.edu/PearlGuterman/Papers/169408/Psychological\_preparedness\_for\_disaster`.

Hamilton, David et al. (2021). “Immersive virtual reality as a pedagogical tool in education: a systematic literature review of quantitative learning outcomes and experimental design”. In: *Journal of Computers in Education* 8.1, pp. 1–32.

Hammarberg, Karin, Maggie Kirkman, and Sheryl de Lacey (2016). “Qualitative research methods: when to use them and how to judge them”. In: *Human reproduction* 31.3, pp. 498–501.

Hanoun, Samer and Saeid Nahavandi (2018). “Current and future methodologies of after action review in simulation-based training”. In: *Proceeding of the 2018 Annual IEEE International Systems Conference (SysCon)*. IEEE, pp. 1–6.

Hattie, John and Shirley Clarke (2018). *Visible learning: feedback*. Routledge.

Herrington, Jan, Thomas C Reeves, and Ron Oliver (2007). “Immersive learning technologies: Realism and online authentic learning”. In: *Journal of Computing in Higher Education* 19.1, pp. 80–99.

Holtzblatt, Karen, Jessamyn Burns Wendell, and Shelley Wood (2004). *Rapid contextual design: a how-to guide to key techniques for user-centered design*. Elsevier.

Homeland Security, Department of (2018). URL: `https://www.dhs.gov/sites/default/files/publications/Virtual\%20Training-Simulation\%20Tool\%20for\%20First\%20Responders.pdf`.

Homeland Security EDGE (2018). *Enhanced Dynamic Geo-Social Environment EDGE | Homeland Security*. (Accessed on 05/30/2018). URL: `https://www.dhs.gov/science-and-technology/EDGE`.

Howie, Scott and Marco Gilardi (2020). “Virtual Observations: a software tool for contextual observation and assessment of user’s actions in virtual reality”. In: *Virtual Reality*, pp. 1–14.

Howie, Scott R and Marco Gilardi (2019). “Virtual Observations of Virtual Reality Simulations”. In: *Proceedings of the 2019 CHI Conference Extended Abstracts on Human Factors in Computing Systems (To Appear)*. CHI EA ’19. Glasgow, Scotland, Uk: ACM.

Howie, Scott Ronald and Marco Gilardi (Mar. 2021). *Methods and Systems for Recording a User Experience*.

Howland, Kate and Judith Good (2015). “Learning to communicate computationally with Flip: A bi-modal programming language for game creation”. In: *Computers & Education* 80, pp. 224–240.

Hsu, Edbert B et al. (2013). “State of virtual reality based disaster preparedness and response training”. In: *PLoS currents* 5.

HTC Vive (2019). *VIVE United Kingdom: Discover Virtual Reality Beyond Imagination*. (Accessed on 18/01/2019). URL: `https://www.vive.com/uk/`.

Hulsman, Robert L and Jane van der Vloodt (2015). “Self-evaluation and peer-feedback of medical students’ communication skills using a web-based video annotation system. Exploring content and specificity”. In: *Patient Education and Counseling* 98.3, pp. 356–363.

Inman, Chris, Vivian H Wright, and Julia A Hartman (2010). “Use of Second Life in K-12 and higher education: A review of research”. In: *Journal of Interactive Online Learning* 9.1, pp. 44–63.

Jensen, Katrine et al. (2017). “Using virtual reality simulation to assess competence in video-assisted thoracoscopic surgery (VATS) lobectomy”. In: *Surgical endoscopy* 31.6, pp. 2520–2528.

Jensen, Lasse and Flemming Konradsen (July 2018). “A review of the use of virtual reality head-mounted displays in education and training”. In: *Education and Information Technologies* 23, pp. 1–15. DOI: 10.1007/s10639-017-9676-0.

Jerald, Jason (2016). *The VR Book: Human-Centered Design for Virtual Reality*. New York, NY, USA: Association for Computing Machinery and Morgan & Claypool. ISBN: 978-1-97000-112-9.

Johnson, Cheryl I and Heather A Priest (2014). “19 The Feedback Principle in Multimedia Learning”. In: *The Cambridge handbook of multimedia learning*, p. 449.

Johnson, Cynthia and Avelino J Gonzalez (2008). “Automated after action review: State-of-the-art review and trends”. In: *The Journal of Defense Modeling and Simulation* 5.2, pp. 108–121.

Jonsson, Anders (2013). “Facilitating productive use of feedback in higher education”. In: *Active learning in higher education* 14.1, pp. 63–76.

Joy Cumming, J and Graham S Maxwell (1999). “Contextualising authentic assessment”. In: *Assessment in education: Principles, policy & practice* 6.2, pp. 177–194.

Kandemir, Cagatay, Omer Soner, and Metin Celik (2018). “Proposing a practical training assessment technique to adopt simulators into marine engineering education”. In: *WMU Journal of Maritime Affairs* 17.1, pp. 1–15.

Kiourt, Chairi, Anestis Koutsoudis, and George Pavlidis (2016). “DynaMus: A fully dynamic 3D virtual museum framework”. In: *Journal of Cultural Heritage* 22, pp. 984–991.

Klomp, Richard W, Josh S Spitalnick, and Dori B Reissman (2011). “Virtual classroom immersion training”. In: *Training Plus Development* 65.1, pp. 38–43.

Kolb, David (Jan. 1984). *Experiential Learning: Experience As The Source Of Learning And Development*. Vol. 1. ISBN: 0132952610.

Konakondla, Sanjay, Reginald Fong, and Clemens M Schirmer (2017). “Simulation training in neurosurgery: advances in education and practice”. In: *Advances in medical education and practice* 8, p. 465.

Krishnan, Rohith R et al. (2019). “A Novel Virtual Reality Game for Disaster Management Applications”. In: *2019 International Conference on Communication and Signal Processing (ICCSP)*. IEEE, pp. 0254–0257.

Kruger, Jessica S and Todd Sage (2020). “Video Grading: Enhancing Clarity and Connection”. In: *Journal of Educational Technology Systems* 48.3, pp. 407–415.

Lahanas, Vasilelios, Evangelos Georgiou, and Constantinos Loukas (2016). “Surgical simulation training systems: box trainers, virtual reality and augmented reality simulators”. In: *International Journal of Advanced Robotics and Automation* 1.2, pp. 1–9.

Lamey, Andy (2015). “Video feedback in philosophy”. In: *Metaphilosophy* 46.4-5, pp. 691–702.

Lang, Partel (2019). *Final IK*. https://assetstore.unity.com/packages/tools/animation/final-ik-14290. [Online; accessed 04-April-2019].

Latham, Kieran et al. (2019). “A Review on the Development of a Virtual Reality Learning Environment for Medical Simulation and Training”. In: *VISUAL 2019*, p. 39.

LaViola, Joseph J et al. (2017). *3D user interfaces*. Addison-Wesley.

Lazar, Jonathan, Jinjuan Heidi Feng, and Harry Hochheiser (2017). *Research methods in human-computer interaction*. Morgan Kaufmann.

Lee, Alfred T (2017). *Flight simulation: virtual environments in aviation*. Routledge.

LinderoEdutainment (2017). *EarthquakeSimulatorVR*. [DIGITAL]. URL: âĂœhttp://store.steampowered.com/app/607590/Earthquake_Simulator_VRâĂİ.

Lopez, Thomas et al. (2017). “A playback tool for reviewing VR experiences”. In: *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. ACM, p. 83.

Lund, Arnold (Jan. 2001). “Measuring Usability with the USE Questionnaire”. In: *Usability and User Experience Newsletter of the STC Usability SIG* 8.

MacCormick, Daniel et al. (2019). “FRVRIT: A Tool for Full Body Virtual Reality Game Evaluation”. In: *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–6.

Magix (2021). *Video Production Software — Unleash Your Creativity*. (Accessed on 01/04/2021). URL: https://www.vegascreativesoftware.com/gb/vegas-pro/.

Mahoney, Paige, Susie Macfarlane, and Rola Ajjawi (2019). “A qualitative synthesis of video feedback in higher education”. In: *Teaching in higher education* 24.2, pp. 157–179.

Malterud, Kirsti, Volkert Dirk Siersma, and Ann Dorrit Guassora (2016). “Sample Size in Qualitative Interview Studies: Guided by Information Power”. In: *Qualitative Health Research* 26.13. PMID: 26613970, pp. 1753–1760. DOI: 10.1177/1049732315617444. eprint: https://doi.org/10.1177/1049732315617444. URL: https://doi.org/10.1177/1049732315617444.

Martins, Luis L and Franz Willi Kellermanns (2004). “A model of business school students’ acceptance of a web-based course management system”. In: *Academy of Management Learning & Education* 3.1, pp. 7–26.

Matthews, Lynnette (2019). “Should we use video technology for giving feedback?” In: *Higher Education Pedagogies* 4.1, pp. 476–487.

Maureen Connolly (2018). *Emergency Management in the Federal Republic of Germany: Preserving its Critical Infrastructures from Hazardous Natural Events and Terrorist Acts*. (Accessed on 06/03/2018). URL: http://training.fema.gov/EMIWeb/edu/Comp.

McCue, Colleen (2007). “2 - Domain Expertise”. In: *Data Mining and Predictive Analysis*. Ed. by Colleen McCue. Burlington: Butterworth-Heinemann, pp. 19–24. ISBN: 978-0-7506-7796-7. DOI: https://doi.org/10.1016/B978-075067796-7/50024-6. URL: https://www.sciencedirect.com/science/article/pii/B9780750677967500246.

McGreevy, James M (2005). “The aviation paradigm and surgical education”. In: *Journal of the American College of Surgeons* 201.1, pp. 110–117.

Meißner, Martin et al. (2017). “Combining virtual reality and mobile eye tracking to provide a naturalistic experimental environment for shopper research”. In: *Journal of Business Research*.

Messick, Samuel (1994). “The interplay of evidence and consequences in the validation of performance assessments”. In: *Educational researcher* 23.2, pp. 13–23.

Microsoft (2021). *C-sharp Programming Guide*. (Accessed on 01/04/2021). URL: https://docs.microsoft.com/en-us/dotnet/csharp/programming-guide.

Miller, Mark Roman et al. (2020). “Personal identifiability of user tracking data during observation of 360-degree VR video”. In: *Scientific Reports* 10.1, pp. 1–10.

Mirchi, Nykan et al. (2020). “The Virtual Operative Assistant: An explainable artificial intelligence tool for simulation-based training in surgery and medicine”. In: *PloS one* 15.2, e0229596.

Mohamadipanah, Hossein et al. (2020). “Can Virtual Reality Be Used to Track Skills Decay During the Research Years?” In: *Journal of Surgical Research* 247, pp. 150–155. ISSN: 0022-4804. DOI: https://doi.org/10.1016/j.jss.2019.10.030. URL: https://www.sciencedirect.com/science/article/pii/S0022480419307607.

Moore, Alec G et al. (2020). “A Formative Evaluation Methodology for VR Training Simulations”. In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, pp. 125–132.

Mossel, Annette et al. (2017). “VROnSite: Towards immersive training of first responder squad leaders in untethered virtual reality”. In: *Virtual Reality (VR), 2017 IEEE*. IEEE, pp. 357–358.

Mueller, Jennifer S, Shimul Melwani, and Jack A Goncalo (2012). “The bias against creativity: Why people desire but reject creative ideas”. In: *Psychological science* 23.1, pp. 13–17.

Murphy, Vanessa et al. (Jan. 2017). ““Keeping It Real”: A Review of the Benefits, Challenges and Steps Towards Implementing Authentic Assessment”. In: *The All Ireland Journal of Teaching and Learning in Higher Education (AISHE-J)* 9.

Newmann, Fred M and Doug A Archbald (1992). “The nature of authentic academic achievement”. In: *Toward a new science of educational testing and assessment*, pp. 71–83.

Nicol, David J and Debra Macfarlane-Dick (2006). “Formative assessment and self-regulated learning: A model and seven principles of good feedback practice”. In: *Studies in higher education* 31.2, pp. 199–218.

Numfu, Manoch, Andreas Riel, and Frédéric Noël (2019). “Virtual reality based digital chain for maintenance training”. In: *Procedia CIRP* 84, pp. 1069–1074.

Nyström, Anita et al. (2014). "Nursing students' experiences of being video-recorded during examination in a fictive emergency care situation". In: *International journal of nursing practice* 20.5, pp. 540–548.

Oculus (2021). *Guidelines for VR Performance Optimization*. (Accessed on 01/04/2021). URL: https://developer.oculus.com/documentation/native/pc/dg-performance-guidelines/.

O'Leary, Michael et al. (2018). "The state-of-the-art in digital technology-based assessment". In: *European Journal of Education* 53.2, pp. 160–175.

Oseni, Zainab et al. (2017). "Video-based feedback as a method for training rural healthcare workers to manage medical emergencies: a pilot study". In: *BMC medical education* 17.1, p. 149.

Pan, Jun J et al. (2015). "Virtual reality training and assessment in laparoscopic rectum surgery". In: *The International Journal of Medical Robotics and Computer Assisted Surgery* 11.2, pp. 194–209.

Patle, Dipesh S et al. (2019). "Operator training simulators in virtual reality environment for process operators: a review". In: *Virtual Reality* 23.3, pp. 293–311.

Pearl, Judea (2014). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier.

Pedram, Shiva et al. (2020). "Investigating the process of mine rescuers' safety training with immersive virtual reality: A structural equation modelling approach". In: *Computers & Education*, p. 103891.

Perry, Ronald W (2004). "Disaster exercise outcomes for professional emergency personnel and citizen volunteers". In: *Journal of Contingencies and Crisis Management* 12.2, pp. 64–75.

Petrie, Helen et al. (2006). "Remote usability evaluations with disabled people". In: *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, pp. 1133–1141.

Pietquin, Olivier, Helen Hastie, et al. (2013). "A survey on metrics for the evaluation of user simulations". In: *Knowledge Engineering Review* 28.1, pp. 59–73.

Qandil, Amjad, Ruba Darweesh, and Abeer Al-Ghananeem (2021). "The Case for Authentic Student Assessment in Distance Digital Pharmacy Education". In: *MedEdPublish* 10.

Ragazzoni, Luca et al. (2015). "Virtual reality simulation training for Ebola deployment". In: *Disaster medicine and public health preparedness* 9.5, pp. 543–546.

Raij, Andrew B and Benjamin C Lok (2008). "Ipsviz: An after-action review tool for human-virtual human experiences". In: *2008 IEEE Virtual Reality Conference*. IEEE, pp. 91–98.

Ramaprasad, Arkalgud (1983). "On the definition of feedback". In: *Behavioral science* 28.1, pp. 4–13.

Raymond, Chad and Simon Usherwood (2013). "Assessment in simulations". In: *Journal of Political Science Education* 9.2, pp. 157–167.

Red Cross (2018). *Earthquake Safety and Survival - Red Cross*. (Accessed on 16/03/2018). URL: http://www.glitchstudios.co/portfolio_page/red-cross/.

Reiners, Torsten, Lincoln C Wood, and Jon Dron (2014). "From chaos towards sense: A learner-centric narrative virtual learning space". In: *Gamification for human factors integration: Social, education, and psychological issues*. IGI Global, pp. 242–258.

Ritter Iii, Kenneth A, Christoph W Borst, and Terrence L Chambers (2015). "Overview and Assessment of Unity Toolkits for Cave Automatic Virtual Environments and Wand Interaction". In: *Int. J. Innov. Educ. Res*, pp. 3–7.

Roberts, Patrick Garfjeld et al. (2017). "Validation of the updated ArthroS simulator: face and construct validity of a passive haptic virtual reality simulator with novel performance metrics". In: *Knee Surgery, Sports Traumatology, Arthroscopy* 25.2, pp. 616–625.

Roy, Elby, Mahmoud M Bakr, and Roy George (2017). "The need for virtual reality simulators in dental education: A review". In: *The Saudi dental journal* 29.2, pp. 41–47.

Sadler, D Royce (1989). "Formative assessment and the design of instructional systems". In: *Instructional science* 18.2, pp. 119–144.

Salgado, Débora Pereira et al. (2018). "A QoE assessment method based on EDA, heart rate and EEG of a virtual reality assistive technology system". In: *Proceedings of the 9th ACM Multimedia Systems Conference*, pp. 517–520.

Sandelowski, Margarete (1986). "The problem of rigor in qualitative research." In: *Advances in nursing science*.

Sankaranarayanan, Ganesh et al. (2018). "Immersive virtual reality-based training improves response in a simulated operating room fire scenario". In: *Surgical endoscopy* 32.8, pp. 3439–3449.

Schloerb, David (Jan. 1995). "A Quantitative Measure of Telepresence". In: *Presence* 4, pp. 64–80. DOI: 10.1162/pres.1995.4.1.64.

Sharma, Ayush et al. (2017). "Virtual reality: blessings and risk assessment". In: *arXiv preprint arXiv:1708.09540*.

Shaw, Emily et al. (2019). "The heat is on: exploring user behaviour in a multisensory virtual environment for fire evacuation". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–13.

Shute, Valerie, Fengfeng Ke, and Lubin Wang (2017). "Assessment and adaptation in games". In: *Instructional techniques to facilitate learning and motivation of serious games*. Springer, pp. 59–78.

Shute, Valerie et al. (2021). "Maximizing learning without sacrificing the fun: Stealth assessment, adaptivity and learning supports in educational games". In: *Journal of Computer Assisted Learning* 37.1, pp. 127–141.

Shute, Valerie J (2008). "Focus on formative feedback". In: *Review of educational research* 78.1, pp. 153–189.

— (2009). "Simply assessment". In: *International Journal of Learning and Media* 1.2, pp. 1–11.

Shute, Valerie J (2011). “Stealth assessment in computer-based games to support learning”. In: *Computer games and instruction* 55.2, pp. 503–524.

Simeone, Adalberto L, Eduardo Velloso, and Hans Gellersen (2015). “Substitutional reality: Using the physical environment to design virtual reality experiences”. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, pp. 3307–3316.

Slater, Mel, Martin Usoh, and Anthony Steed (1995). “Taking steps: the influence of a walking technique on presence in virtual reality”. In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 2.3, pp. 201–219.

Slater, Mel and Sylvia Wilbur (Dec. 1997). “A Framework for Immersive Virtual Environments Five: Speculations on the Role of Presence in Virtual Environments”. In: *Presence: Teleoper. Virtual Environ.* 6.6, 603–616. ISSN: 1054-7460. DOI: `10.1162/pres.1997.6.6.603`. URL: `https://doi.org/10.1162/pres.1997.6.6.603`.

Slater, Mel et al. (Sept. 2009). “How we experience immersive virtual environments: The concept of presence and its measurement”. In: *Anuario de Psicologia* 40.

Stone, Robert, Timothy Snell, and Neil Cooke (2016). “An inexpensive underwater mine countermeasures simulator with real-time 3D after action review”. In: *Defence technology* 12.5, pp. 367–379.

Sweller, John, Jeroen JG Van Merrienboer, and Fred GWC Paas (1998). “Cognitive architecture and instructional design”. In: *Educational psychology review* 10.3, pp. 251–296.

Takala, Tuukka M (2014). “RUIS: A toolkit for developing virtual reality applications with spatial interaction”. In: *Proceedings of the 2nd ACM symposium on Spatial user interaction*. ACM, pp. 94–103.

Thanyadit, Santawat, Parinya Punpongsanon, and Ting-Chuen Pong (2020). “Demonstration of ObserVAR: Visualization System for Observing Virtual Reality Users Using Augmented Reality”. In: *SIGGRAPH Asia 2020 XR*. SA ’20. Virtual Event, Republic of Korea: Association for Computing Machinery. ISBN: 9781450381116. DOI: `10.1145/3415256.3421495`. URL: `https://doi.org/10.1145/3415256.3421495`.

Thomsen, Ann Sofia Skou et al. (2017). “High correlation between performance on a virtual-reality simulator and real-life cataract surgery”. In: *Acta ophthalmologica* 95.3, pp. 307–311.

Thurston, Casey L and Glenn A Martin (2011). “A generic after action review capability for game-based training”. In: *International Conference on Engineering Psychology and Cognitive Ergonomics*. Springer, pp. 628–634.

Todi, Kashyap et al. (2018). “Familiarisation: Restructuring layouts with visual learning models”. In: *23rd International Conference on Intelligent User Interfaces*, pp. 547–558.

Tromp, Jolanda G, Anthony Steed, and John R Wilson (2003). “Systematic usability evaluation and design issues for collaborative virtual environments”. In: *Presence: Teleoperators & Virtual Environments* 12.3, pp. 241–267.

Unity (2021a). *Configuring Animation Clip Recorders: Unity Recorder: 2.1.0-preview.1*. (Accessed on 18/01/2021). URL: https://docs.unity3d.com/Packages/com.unity.recorder@2.1/manual/RecorderAnimation.html.

— (2021b). *Order of execution for event functions*. (Accessed on 03/04/2021). URL: https://docs.unity3d.com/Manual/ExecutionOrder.html.

— (2021c). *ScriptReference*. (Accessed on 18/01/2021). URL: https://docs.unity3d.com/2018.3/Documentation/ScriptReference/Animations.GameObjectRecorder.html.

Upadhyay, Ashwani Kumar and Komal Khandelwal (2019). "Artificial intelligence-based training learning from application". In: *Development and Learning in Organizations: An International Journal*.

Valenzeno, Laura, Martha W Alibali, and Roberta Klatzky (2003). "Teachers' gestures facilitate students' learning: A lesson in symmetry". In: *Contemporary Educational Psychology* 28.2, pp. 187–204.

Valve (2021). *Steam Storefront*. (Accessed on 18/01/2021). URL: https://store.steampowered.com/.

Veterans Affairs (2018). *Virtual reality boosts job-interview skills for Veterans with PTSD*. (Accessed on 05/30/2018). URL: https://www.research.va.gov/currents/0715-6.cfm.

Vincent, Dale S et al. (2008). "Teaching mass casualty triage skills using immersive three-dimensional virtual reality". In: *Academic Emergency Medicine* 15.11, pp. 1160–1165.

Virtual Medical Center (2018). *VA-VMC*.

Von Spiczak, J et al. (2007). "Multimodal event streams for virtual reality". In: *Multimedia Computing and Networking 2007*. Vol. 6504. International Society for Optics and Photonics, p. 65040M.

Vos, Lynn (2015). "Simulation games in business and marketing education: How educators assess student learning from simulations". In: *The International Journal of Management Education* 13.1, pp. 57–74.

VRM Switzerland (2021). *Professional Flight Training Solutions*. (Accessed on 29/04/2021). URL: https://vrm-switzerland.ch/.

Wan, Zeying and Yulin Fang (2006). "The role of information technology in technology-mediated learning: A review of the past for the future". In: *AMCIS 2006 Proceedings*, p. 253.

Wang, Minhong (2018). "Learning goals and performance measurement in the workplace". In: *E-Learning in the Workplace*. Springer, pp. 85–92.

Wiggins, Grant (1990). "The case for authentic assessment". In: *Practical assessment, research, and evaluation* 2.1, p. 2.

Wijewickrema, Sudanthi et al. (2017). "Feedback Techniques in Computer-Based Simulation Training: A Survey". In: *arXiv preprint arXiv:1705.04683*.

Wismer, Philip et al. (2021). "Immersive virtual reality as a competitive training strategy for the biopharma industry". In: *Nature Biotechnology* 39.1, pp. 116–119.

Wood, Lincoln et al. (2013). "The role of gamification and game-based learning in authentic assessment within virtual environments". In: *Research and development in higher education: The place of learning and teaching*, pp. 514–523.

Xie, Jiacheng et al. (2018). "A remote VR operation system for a fully mechanised coal-mining face using real-time data and collaborative network technology". In: *Mining Technology* 127.4, pp. 230–240.

Yin, Robert K (2015). *Qualitative research from start to finish*. Guilford Publications.

Yoganathan, S et al. (2018). "360 virtual reality video for the acquisition of knot tying skills: a randomised controlled trial". In: *International Journal of Surgery* 54, pp. 24–27.

Appendix A

Evaluation Semi-Structured Interview Questions and Questionnaires

Semi-Structured Interview Questions

Questions will be developed based on observations made by the lead investigator and comments made by the participant during their session. This is an example of the type of questions that will be asked:

1. What **features** did you **like** about the observation Method [*identified*]?
2. Thinking back to the [*identified*] observation Method. What was the most **important feature** for assessment of the trainee?
3. Thinking back to the [*identified*] observation Method. What was the **most effective feature** for assessment of the assessment of the trainee?
4. What made you choose that feature for being most effective?
5. Thinking back to the [*identified*] observation Method. Was there anything you were unable to do using that Method that you wish you could have?
6. [Question for Participant] Do you have anything you want to say or any opinions you would like to voice?

Identified Observation methods will be used to reflect the question to findings from the questionnaire.

General Questions

Introduction & Aims:

The following questionnaire will consist of five sections that regard the method of observation you have just used to assess and provide feedback for a Virtual Reality training simulation dataset. If you are unsure of the method previously used, please ask the investigator.

The five sections will determine different factors of the method used with regards to Usefulness and Usability. Overall, the questionnaire will provide insight into the application of the observational method to complete the objective of assessing and providing feedback for VR training simulations.

* Required

1. Previous Method Used: (please ask investigator if unsure) *

Mark only one oval.

- Data
- Video
- Virtual Observation

2. How confident do you think the trainee was during their assessment? (Dataset A) *

Mark only one oval.

- Extremly Confident
- Confident
- Neither Confident or Non-Confident
- Non-Confident
- Extremly Non-Confident
- I am unsure of the participants confidence level

3. How confident do you think the trainee was during their assessment? (Dataset B) *

Mark only one oval.

- Extremly Confident
- Confident
- Neither Confident or Non-Confident
- Non-Confident
- Extremly Non-Confident
- I am unsure of the participants confidence level

Analysis of Observing VR Simulations (Usefulness)

Introduction & Aims:
The following two sections will involve analysis of the usefulness of the method for observation. The aim of these two sections is to clarify how useful the method was to achieving tasks and the overall objective for assessment and feedback of VR training simulations - Part 1.

4. Please Indicate your level of agreement with the following statements: *

Mark only one oval per row.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I acquired all information I sought from the trainee VR assessment using this method of observation.	⬭	⬭	⬭	⬭	⬭
I think the data I acquired from this Observation method enabled suitable analysis of the VR trainee assessment.	⬭	⬭	⬭	⬭	⬭
I think the information I acquired from this method of Observation was insightful of the VR trainee assessment.	⬭	⬭	⬭	⬭	⬭
I was able to understand what the VR trainee was doing during the	⬭	⬭	⬭	⬭	⬭

simulation.					
I found this method of observation effective for reviewing VR trainee performance.	◯	◯	◯	◯	◯
I found this method of observation effective for generating feedback for the VR trainee.	◯	◯	◯	◯	◯

(Usefulness) Questionnaire Part - 2

Introduction & Aims
The following two sections will involve analysis of the usefulness of the method for observation. The aim of these two sections is to clarify how useful the method was to achieving tasks and the overall objective for assessment and feedback of VR training simulations - Part 2.

5. Please Indicate your level of agreement with the following statements:

Mark only one oval per row.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
The method of observation I used was effective.	◯	◯	◯	◯	◯
The method of observation I used helped my observation be more productive.	◯	◯	◯	◯	◯
The method of observation I used was useful for observing VR simulations.	◯	◯	◯	◯	◯
The method of observation I used gave me control over my observations.	◯	◯	◯	◯	◯
The method of observation I used made achievingmy goals easier.	◯	◯	◯	◯	◯
The method of observation I	◯	◯	◯	◯	◯

used met the needs I have during assessment of trainees.					
The method of observation I used met my expectations.	◯	◯	◯	◯	◯

Analysis of Observing VR Simulations (Usability - Part 1)

Introduction & Aims: (PART 1)
The following two sections will involve analysis regarding the usability of the observational method. The aim of these two sections is to determine the usability of the method for achieving goals and objectives. Overall, the questionnaire is designed to discover the practical use of the method and its features - Part 1.

6. Please Indicate your level of agreement with the following statements:

Mark only one oval per row.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
The method of observation I used was time-efficient for analysing VR trainee assessment.	⬭	⬭	⬭	⬭	⬭
The method of observation I used was time-efficient for providing feedback to trainees.	⬭	⬭	⬭	⬭	⬭
The method of observation I used made providing feedback with evidence to the VR trainee easy.	⬭	⬭	⬭	⬭	⬭
I did not experience any frustration while using this observation method.	⬭	⬭	⬭	⬭	⬭

(Usability - Part 2)

Introduction & Aims: (PART 2)
The following two sections will involve analysis regarding the usability of the observational method. The aim of these two sections is to determine the usability of the method for achieving goals and objectives. Overall, the questionnaire is designed to discover the practical use of the method and its features - Part 2.

7. Please Indicate your level of agreement with the following statements:

Mark only one oval per row.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
If VR was adopted by the company, I would use this system frequently in the future to assess VR simulations.	◯	◯	◯	◯	◯
I found the system unnecessarily complex.	◯	◯	◯	◯	◯
I thought the system was easy to use.	◯	◯	◯	◯	◯
I think I would need support to be able to use this system.	◯	◯	◯	◯	◯
I found the various functions and functionality of the system were well integrated.	◯	◯	◯	◯	◯
I learned how to use this system very quickly.	◯	◯	◯	◯	◯
I found this method of observation	◯	◯	◯	◯	◯

very cumbersome to use.					
I felt very confident using this method of observation.	⬭	⬭	⬭	⬭	⬭
I needed to learn a lot of things before I could get going with this method of observation.	⬭	⬭	⬭	⬭	⬭
I found there was too much inconsistency in this method of observation.	⬭	⬭	⬭	⬭	⬭

8. Overall, I would rate the user-friendliness of this method of observation as

Mark only one oval.

- ⬭ Awful
- ⬭ Poor
- ⬭ OK
- ⬭ Good
- ⬭ Excellent

Google Forms

Usefulness of Virtual Observation Features

Introduction & Aims:
The following questionnaire will reflect upon your experience using the Virtual Observation software and its usefulness for assessing and providing feedback for Virtual Reality training simulations.

The questionnaire aims to identify which of the following features are considered useful for the purpose of observing, assessing and providing feedback of Virtual Reality training simulations.

Please rank from 1 (absolutely not Useful) to 5 (Extremly Useful) the USEFULNESS of each of the following features of Virtual Observation:

(Modified Yaelle Overview Questions)
* Required

1. Please Indicate your level of usefulness with the following functionalities: *

Mark only one oval per row.

	Absolutely not Useful	Slightly Useful	Moderately Useful	Very Useful	Extremly Useful
Ability to control simulation time for Pauseing or Rewinding the observed VR simulation.	⬭	⬭	⬭	⬭	⬭
Ability to skip to different times in the VR simulation.	⬭	⬭	⬭	⬭	⬭
The Time and Date information of when the VR trainee operated the simulation.	⬭	⬭	⬭	⬭	⬭
How long the user VR Simulation lasted.	⬭	⬭	⬭	⬭	⬭
The Timeline of actions and events conducted by the VR trainee throughout their assessment.	⬭	⬭	⬭	⬭	⬭
A Timeline of objectives completed or failed	⬭	⬭	⬭	⬭	⬭

throughout their assessment.					
An estimated skeleton posture of the VR trainee in real-life.	◯	◯	◯	◯	◯
Voice-Recording of the VR trainee throughout the VR assessment.	◯	◯	◯	◯	◯
Ability to draw/highlight ontop of the VR simulation for providing feedback.	◯	◯	◯	◯	◯
Ability to capture images of the observation.	◯	◯	◯	◯	◯
Ability to record videos of the the observation.	◯	◯	◯	◯	◯
An overall summary of key simulation information, such as objectives passed or failed by the VR trainee.	◯	◯	◯	◯	◯

Ranking Observational Methods

Introduction & Aims:
The following questionnaire will reflect upon your experience using all three methods of observation (Data, Video and Virtual Observation) for assessing and providing feedback for Virtual Reality simulations. The aims of this questionnaire is to determine the preference of observational methods within individual contexts.

Please rank each method of observation in your preferance from First to Third Choice.

This questionnaire will only be conducted after the participant has reviewed the data-set(s) using ALL three methods of observation)

* Required

1. Please rank from First to Third, in your opinion, what method of observation was most effective for observing the VR trainee assesment. *

Mark only one oval per row.

	First Choice	Second Choice	Third Choice
Virtual Observation	⬭	⬭	⬭
Video	⬭	⬭	⬭
Data	⬭	⬭	⬭

2. Please rank from First to Third, in your opinion, what observation method provided the best time-efficent way for observing the VR trainee assessment. *

Mark only one oval per row.

	First Choice	Second Choice	Third Choice
Virtual Observation	⬭	⬭	⬭
Video	⬭	⬭	⬭
Data	⬭	⬭	⬭

3. Please rank from First to Third, in your opinion, what methods of observation were easiest to use. *

Mark only one oval per row.

	First Choice	Second Choice	Third Choice
Virtual Observation	⬭	⬭	⬭
Video	⬭	⬭	⬭
Data	⬭	⬭	⬭

4. Please rank from First to Third, in your opinion, what methods best presented the VR user's simulation data for analysis. *

Mark only one oval per row.

	First Choice	Second Choice	Third Choice
Virtual Observation	◯	◯	◯
Video	◯	◯	◯
Data	◯	◯	◯

5. Please rank from First to Third, in your opinion, what methods best allowed you to provide feedback to the VR trainee. *

Mark only one oval per row.

	First Choice	Second Choice	Third Choice
Virtual Observation	◯	◯	◯
Video	◯	◯	◯
Data	◯	◯	◯

6. Please rank from First to Third, in your opinion, what observation method provided the best time-efficent way for providing feedback to the VR trainee on their performance. *

Mark only one oval per row.

	First Choice	Second Choice	Third Choice
Virtual Observation	⬭	⬭	⬭
Video	⬭	⬭	⬭
Data	⬭	⬭	⬭

7. Please rank from First to Third, in your overall opinion what order you would select to use in the future for observation of VR trainee assessment. *

Mark only one oval per row.

	First Choice	Second Choice	Third Choice
Virtual Observation	⬭	⬭	⬭
Video	⬭	⬭	⬭
Data	⬭	⬭	⬭

Appendix B

Ethics Forms

Information Sheet: Observation of Virtual Reality Simulations

Thank you for agreeing to participate in this research. As part of my PhD project I am investigating how Virtual Reality simulations can be assessed to provide feedback to VR users.

During this study you will use several methods of observation to assess VR user experiences and do your best to provide them with relevant feedback on their performance. At the same time, you will be asked to fill in questionnaires based on your feelings of the observation methods, which will be followed by a short commentary session of your experience and a semi-structured interview.

The objective this session is to gather data that will allow determine to advantages and disadvantages of observational methods, using four different factors for comparison. The findings from this study will be used in my PhD thesis.

Interviews and the discussions will be recorded using an audio recording device. These recordings will be transcribed to written format to make them anonymous. During your participation using the four methods of observation, the screen of the computer will be recorded. This will video capture the computer display from the monitor. Eye-tracking will also be used to monitor your gaze while observing VR user experiences. Video, Audio and Eye-Tracking gaze will be combined during analysis to create a detailed visualisation of your observation.

Should you no longer wish to participate in the study, you are able to withdraw your participation without the need to give a justification. If you wish to stop the interview at any time for any reason, please indicate to the lead investigator that you no longer wish to continue.

You will have up to a month after the study has been conducted to revoke your consent for participation. You can do this by emailing Scott Howie (Scott.Howie@uws.ac.uk) with your unique identifier key that will be provided to you on your debrief sheet.

If you have any questions or concerns please feel free to ask the investigator at any time before, during or after the experiment.

Comparison Methods of Observation for Virtual Reality

Consent Form

By participating in this research, you give consent to the following: Please Initial Each Box

1. I confirm that I have read and understood the information sheet for this study. I have had the opportunity to consider the information.
2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason.
3. I agree to have this interview recorded with an audio recording device. I also understand that the audio data will be transcribed.
4. I understand that data collected during the study, may be looked at by researchers and collaborators taking part in this research. I give permission for these individuals to have access to the data collected.
5. I understand that I have agreed to have my eye-gaze tracked and recorded while using the methods of observation.
6. I understand that I have agreed to have my participation using the methods of observation recorded virtually using a screen-recording program on the PC (Computer screen will be recorded)
7. I agree that data collected in this study can be used in follow-up research of the same kind.
8. I understand that I will have up-to submission of this research as a PhD thesis to withdraw my consent by contacting Scott Howie (Scott.Howie@uws.ac.uk). After this time has elapsed, data will be analysed and submitted for publication, and it won't be possible to remove or withdrawn them.

Date ________________

Signature_________________________

UNIVERSITY OF THE WEST *of* SCOTLAND

14/04/2018

Dear Scott Howie,

Your application 4754: Analysing the Spatial Mapping Potential of Different Locomotion Methods in Virtual Reality, submission 3495, has been approved by the Engineering and Computing SEC . You may now proceed with your study. If you wish to make any significant changes to the study you must seek the committee's approval before actioning them.

Good luck with your research.

Prof Malcolm Crowe

UNIVERSITY OF THE WEST of SCOTLAND

UWS

06/11/2019

Dear Scott Howie,

Your application 9333: Swagelok PhD Collaboration – Assessment of Virtual Reality Simulations, submission reference 7365, has been approved, with conditions, by the Computing, Engineering and Physical Sciences SEC. **You may proceed with your study provided you meet the conditions outlined below**;

Title	Comment
Please upload a copy of your Consent Form(s)	On the consent form, it would be an improvement to include a tick box for each item, that the participants can acknowldege their specific consent to each point, and potentially capture a specific unwillingness, for example perhaps to being recorded.

If you wish to make any significant changes to your study you must seek the committee's approval before actioning them.

Good luck with your research.

Dr John Hughes

UNIVERSITY OF THE WEST of SCOTLAND UWS

29/09/2020

Dear Scott Howie,

Your application 12065: Assessment of Virtual Reality Simulations, submission reference 10584, has been approved, with conditions, by the Computing, Engineering and Physical Sciences SEC. **You may proceed with your study provided you meet the conditions outlined below**;

- Please note the request to extend the withdrawal time limeit to a flexible limit related to final publication, submission for publication or submission of thesis.

Title	Comment
Please give a full summary of the purpose, justification, design and methodology of the planned study: (word limit 1000 words)	The final point about the one month limit on withdrawal- please change this to the possibility of withdrawal up to the point of publication/submission of your dissertation. Participants should be able to withdraw at any point in the process, until changes can literally not be made.
Please upload a copy of the Questionnaire or Interview Schedule	Presumably the formatting of the questions will be improved in the online version of the forms? If not please try to make the options on the left column (when it is a longer statement) easier to read- probably by widening the column.
Please upload a copy of the Participant Information Sheet(s)	Same point as earlier about time limit for withdrawal- please change to earlier time limit at point of publication or submission of thesis.
Please upload a copy of your Consent Form(s)	Good consent form- very nice to see itemisation of individual consent items. However, the point no.4 on "..be looked at by researchers and collaborators taking part in this research"- this is vague on who these people are. Can you be more specific as to the roles of these people and the organisations they work for, please? Again- the same point to be made on withdrawal time limit- please change No 8.

If you wish to make any significant changes to your study you must seek the committee's approval before actioning them.

Good luck with your research.

Dr John Hughes

Appendix C

Guidelines for Judging Participant Competency

UWS

OIL-RIG HAZARD IDENTIFICATION

GUIDELINES FOR PARTICIPANT COMPETENCY

SCOTT HOWIE

Contents

Understanding the Steps

The document includes the step-by-step guidelines for the trainee to safely find and identify the fault in the line.

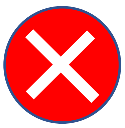
Steps marked with **Critical Safety Failure** must be completed safely while abiding by the requirements laid out at the step. If the **Critical Safety Failure** is triggered, the participant has conducted an action which is unsafe, causing the simulation to be immediately stopped.

Failure Guidelines

There are two failure options within the simulation, `**Minor Error(s)**’ and `**Critical Safety Failure**‘. A minor error is an incorrect action or response, but is not considered a safety critical issues, meaning the error is not dangerous. These minor errors should be corrected during feedback. `**Critical Safety Failure**‘ errors are unsafe mistakes which will or may result in unsafe or dangerous consequences to the trainee, surrounding or operation of the oil-rig.

Minor and Critical Safety Failure errors should be corrected by feedback to notify and correct the user of their error.

`**Minor Error(s)**’ are when the user has missed or incorrectly completed an action. These steps are not critical to the safety of the operation but should be identified and corrected during feedback.



`**Critical Safety Failure**‘ errors are when the trainee does something unsafe, such as attempting vent the line or open the fitting while the line is still pressurised. This icon is shown during steps which have potential for the participant to conduct an unsafe action.

1. Clearing Area of Hazards

Before beginning work at the fault area, the work area must be cleared of hazards. The hazards within the simulation include A Bucket, Step-Stool and Scattered Left-Over Pipes. All these hazards are located on the work area floor and can be cleared in any order.

Figure 1- The three hazards (Step-Stool, Bucket and Pipes).

2. Applied Work Barrier

Once given the affirmative from the systems manager, the participant will be asked to close off the working area by using barrier to cordon of the stairwell, as seen in Figure 2.

Figure 2 - Barrier setup Correctly

3. Identified Problem with Snoop Detector

Using the snoop detector, the participant will look for the fault which is located at the second vent gauge and vent control release. When the snoop detector liquid is applied to a fault area, the pressure leak in the pipe fitting will generate bubbles.

Figure 3 - Area of Fault.

Figure 4 - When the fault is found by the snoop detector, bubbles are generated, signifying the leak.

4. Reported Identification of Fault

When the fault is detected, the trainee should inform the system manager with the following response:

```
"I've found the fault, I need to isolate PI1234 and investigate the leak. Am I ok to proceed?"
```

Other responses are incorrect, forcing the trainee to respond with an alternative option.

Figure 5 - The trainee should report the fault to the system manager using Option 2: "I've found the fault, I need to isolate PI1234 and investigate the leak. Am I ok to proceed?"

If the trainee response is not correct <u>during their initial reply</u>, the error should be considered a Minor Error.

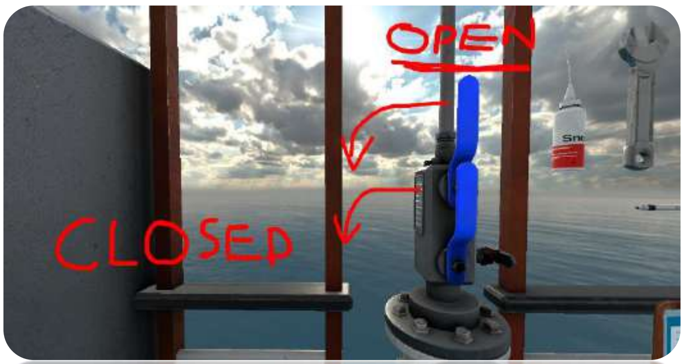
5. Tagged Fault Area

Before attempting to resolve the fault, the trainee should apple an Intevention tag the area to indicate that the pipleline is under maintenance.

Figure 6 - The inspection tag should be fitted before the pipe is disassembled.

6. Shut-Off Pressure Levers in Correct Order

CRITICAL SAFETY FAILURE: Both Pressure Levers must be in the closed position prior to venting the line. Only when BOTH Pressure Levers are in the closed position is the line safe to vent. The Pressure Levers are in the open position when facing upwards as demonstrated in Figure 7. The Pressure Levers are in the closed position when they are down as indicated in Figure 8.



The trainee should operate both Pressure Levers, moving them onto the closed position before conducting any further work on the line. The Pressure Levers are considered open unless BOTH levers are in the closed position as seen in Figure 8.

Figure 7 - Pressure Levers are in OPEN position.

Figure 8 - Pressure Levers are in CLOSED position.

7. Vented Pressure

CRITICAL SAFETY FAILURE: If the system has not been isolated by closing both Pressure Levers (Step 6), with BOTH levers in the CLOSED position, venting the system at this step will blow the line. Refer to Step 6, `Shut-Off Pressure Levers in Correct Order' for more information.

The system can be vented using either the Left OR Right vent levers as seen in Figure 9 and 10. Venting the system depressurises the line, making it safe to dismantle.

Figure 9 demonstrates the system pressured before being vented, while Figure 10 shows the system after it has been safely vented and is not depressurized, ready to be disassembled.

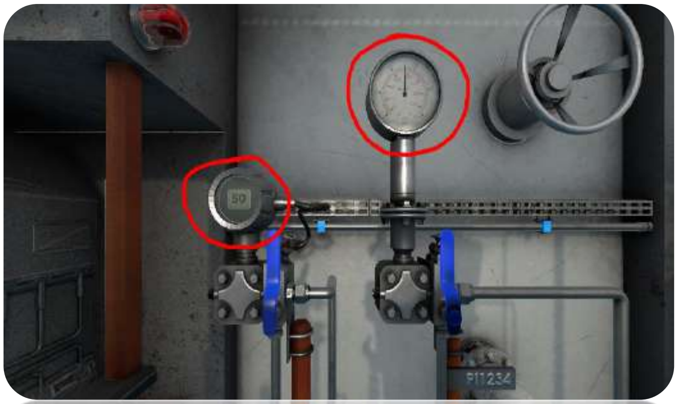


Figure 9 - The pressure gauges are NOT zero, signifying the line is pressurised.

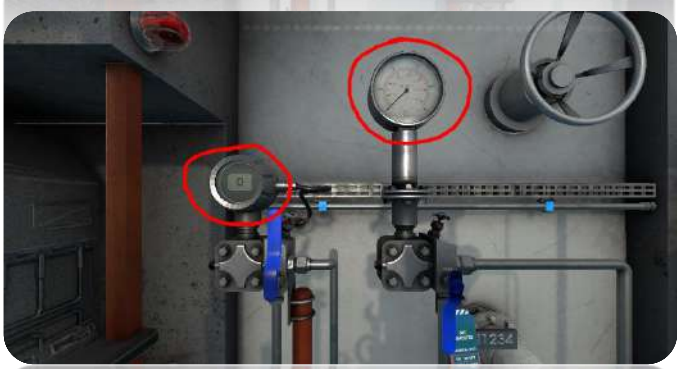
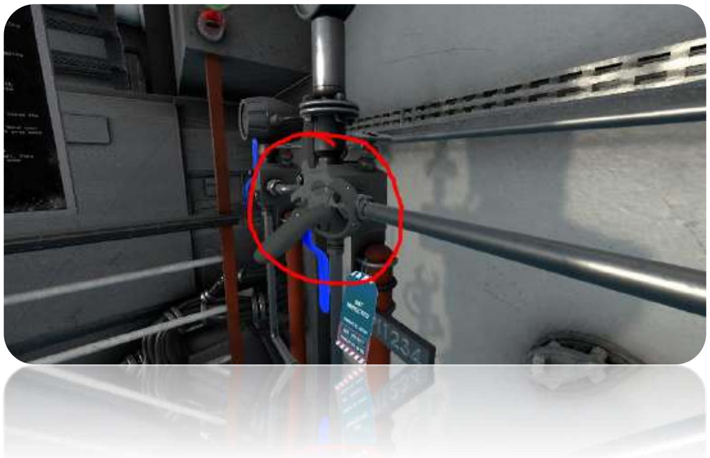


Figure 10 - The pressure gauges are zero, signifying the line is depressurised.

8. Measured with Gap Inspection Gauge



Before dismantling the pipe, the trainee should inspect the gap of the nut and bolt using the inspection gauge. This will signify to the trainee

Figure 11 - The use of the inspection gauge against the pipe fitting, confirming the nut is not the source of the fault.

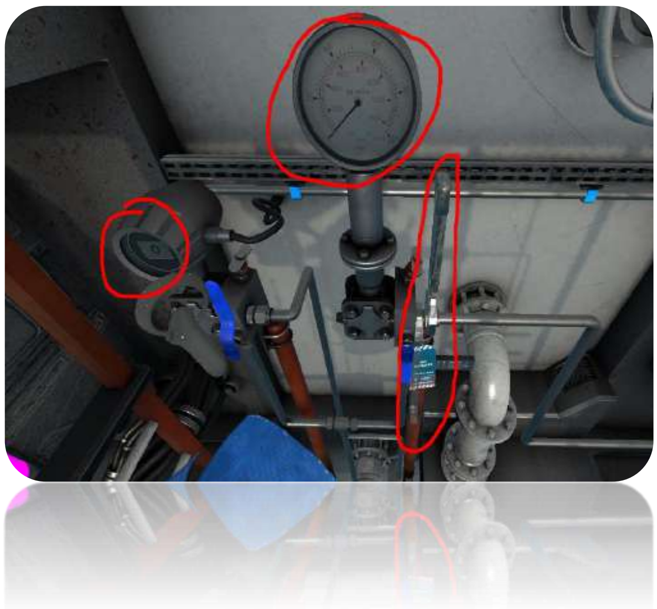
9. Marked Fitting Body, Nut & Pipe on Fault Area

The pen should be used to mark the Body, Nut and Pipe that is being dismantled as seen in Figure 12.

Figure 12 - Pen markings indicating the areas being dismantled.

10. Disassembled the Pipe Fitting

CRITICAL SAFETY FAILURE: If the system has not been isolated and vented, disassembling the pipe fitting will blow the line. Refer to Step 7, `Vented Pressure' for more information.



After the line has been depressurised as indicated by both pressure gauges reading 0, as viewable in Figure 13. The participant should connect the spanner to the nuts and bolts and once confirming the line is depressurized, operate the spanner to dismantle the pipe fitting. This can only safely be conducted when BOTH Isolation Levers are in the CLOSED position (Step 6) and the system has been vented (Step 7).

Figure 13 - The spanner attached to the system after the line has been depressurised.

11. Identified Correct Ferrule Fault

The participant is expected to identify that the ferrule fault has debris attached, making it unfit for purpose. This fault should be reported as the main fault to the system manager via the radio.

Figure 14- The Ferrule and source of the fault. The trainee should inform the System Manager that the fault is: "There is debris on the front ferrule" via the radio.

Trainee correct response:

"There is debris on the front ferrule"

If the trainee response is not correct, the mistake should be considered a Minor Error.

12. Identified Correct Additional Hazard

At the end of the simulation, the system manager will contact the trainee via the radio and ask them to look around the work site to see if they can find any other hazards. The trainee should respond to the system manager that a pressure gauge is unsafely connected to the system and is not adequately supported.

Figure 15 - The trainee should identify the pipe gauge is not adequately attached to a work surface and notify the System Manager when prompted: "There is a heavy pressure gauge on the end of a thin tube above the fault area".

Trainee correct response:

"There is a heavy pressure gauge on the end of a thin tube above the fault area".

If the trainee response is not correct, the mistake should be considered a Minor Error.

Appendix D

Industrial Training Assessors Requirement Collection Study: Semi-Structured Interview Questions

Swagelok Instructor Interview

The goal of these interview questions is to enable me to develop software for Virtual Reality training simulations so that you can assess and give feedback of trainee performances in equal or superior methods that you currently use. In this interview, I will ask you a set of questions based on the understanding I have interpreted training and feedback methods. If you can think of information that you believe would aid my research objective, please feel free to talk at length and control the structure of the interview. The questions I have are only a rough guide.

Part 1: Demographic Information

Question 1: Time of qualification as an Instructor?

Prompt 1a: How long have you been assessing training?

Question 2: Can you describe briefly your background and experience teaching and training?

Prompt 2a: How many training and testing scenarios do you estimate you have monitored?
Prompt 2b: What types of training scenarios do you monitor?
Prompt 2c: How do you monitor training and assessments?
Prompt 2d: Why do you monitor trainees using this method?
Prompt 2e: How long do you estimate a single training and/or assessment scenario takes?

Part 2 Interview Questions:

Question 1: Can you explain in as much detail as possible your last training or teaching session, including trainees, tasks, objectives, along with trainee feedback and learning results?

Prompt 1a: How many people were you training and assessing?
Prompt 1b: How long did this individual module last?
Prompt 1c: How much time is allocated to each individual trainee for training and individual feedback?
Prompt 1d: How do you keep track of monitoring all trainees at the same time?
Prompt 1e: Do you think you may have missed errors conducted by trainees? How often do estimate you may have missed? Can you explain why you may have missed these errors?

Question 2: As part of my research into assessment of training, I am hoping to understand from both the instructor and trainee perspectives regarding feedback handling for training and assessment (or testing). Can you describe in as much detail the process for how you provide feedback to trainees?

Prompt 2a: How to you correct errors and misunderstandings that trainees have?
Prompt 2b: Have you ever experienced difficulty when attempting to teach trainees who are struggling to learn?

Prompt 2ba: How did you resolve these difficulties?

Prompt 2c: How many people observe trainees training – The ratio from instructors to trainees?

Prompt 2ca: Do you think this is a suitable ratio of instructors to trainees?

Prompt 2d: Can you describe an experience when you had difficulty observing trainees during training or testing?

Prompt 2da: Did you resolve this difficulty?

Prompt 2db: How did you resolve this difficulty? / How would you have resolved this difficulty on reflection?

Question 3: What do you look for when observing training performance?

Prompt 3a: Are you interested in the whole training scenario or do you focus on selective elements of the trainee's performance?

Prompt 3b: When observing a trainee, what do you focus your attention on?

Prompt 3c: Do you correct trainee mistakes instantly during training or after when training has concluded?

Prompt 3d: How often do you think you have may have missed an error conducted by the trainee?

Prompt 3e: What happens if you are unsure about the correctness of a trainee's performance? Do you collaborate with other trainers?

Question 4: Can you describe in detail what happens after training and assessment scenarios are complete?

Prompt 4a: Do they receive a form that details their performance? [Training/Assessment]

Prompt 4aa: Do they take anything away with them such as feedback form or performance evaluation?

Prompt 4ab: What does this form contain?

Prompt 4b: How is their performance logged for future reference?

Prompt 4ba: What is stored?

Prompt 4c: How long does a trainee have to wait until the receive feedback or be told if they have passed or failed the test?

Prompt 4d: In situations that the trainee does not accept the outcome of their training or testing performance due to an error they have committed; how do you prove they did commit the error?

Prompt 4da: In situations a trainee has committed an error, is it possible that you may miss the error being conducted, and notice the error at a later stage of the training or assessment? (For example, the user has used an incorrect hose, but the fault is not noticed while the trainee is still on that task).

Question 5: For an ideal scenario, if you were not bound by any financial or physical constraints, what would you implement to improve your observation and assessment abilities for reviewing training or testing?

Prompt 5a: Would you video record trainees?

Prompt 5aa: What would you do with the recordings?

Prompt 5ab: Would you be able to use video recordings to assess trainee performance?

Prompt 5b: How would you improve your method of observing trainees?

Prompt 3ba: For example, would you change training location or observe training using alternative methods?

Prompt 5c: How you would improve your method of marking trainees' performances?

Prompt 3ca: Not bound any physical or financial constraints, what would you do to improve your ability to mark and judge trainee performance?

Prompt 5d: How would you improve the feedback trainees receive?

Prompt 3ca: Not bound any physical or financial constraints, what would you do to improve your ability to give useful individual feedback on trainee performance?

Prompt 5e: Can you think of any other equipment or software which you would like to see implemented to improve training assessment or trainee feedback

Question 6: Imagine in a scenario you were only able to observe the assessment of a trainee live (in real-time) using a camera which you could control the position and rotation of as required. Can you describe how you would approach using this method to review trainee performance as close as possible to how you currently conduct person in person?

Prompt 6a: What if this was a pre-recorded assessment that is not live, but you could still control the position and rotation of the camera?

Prompt 6b: Assuming you had complete control of the recording, would you fast-forward, rewind or skip parts of the assessment?

Prompt 6ba: How would you use these fast-forwarding/skipping/rewinding tools to help you assess trainee performance?

Appendix E

Thematic Maps From Evaluation

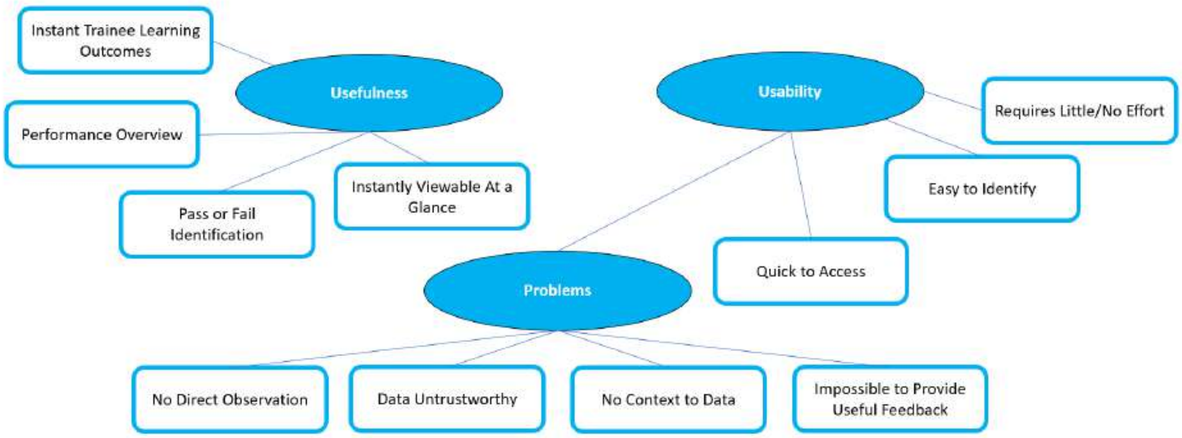


Figure E.1: Initial Analytical Data Thematic Map.

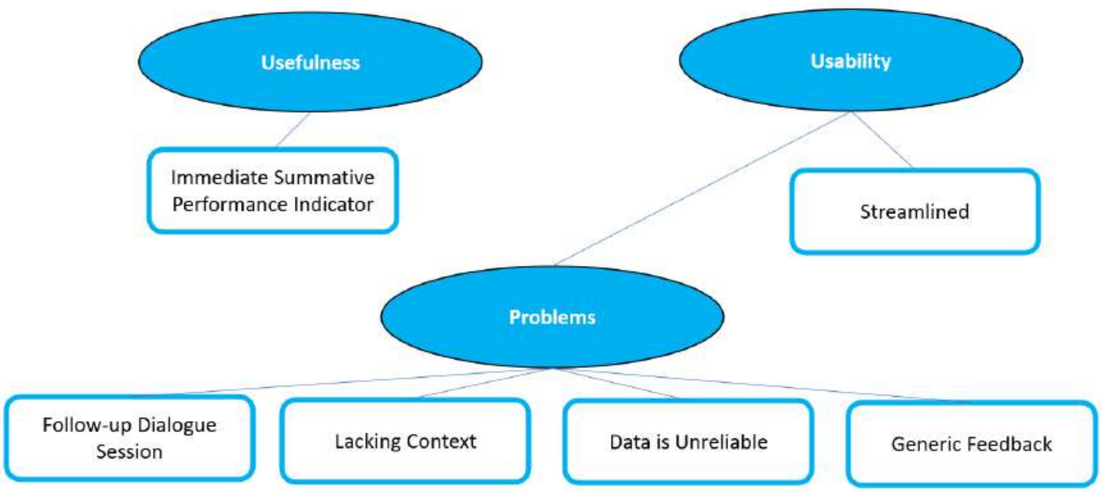


Figure E.2: Final Analytical Data Thematic Map.

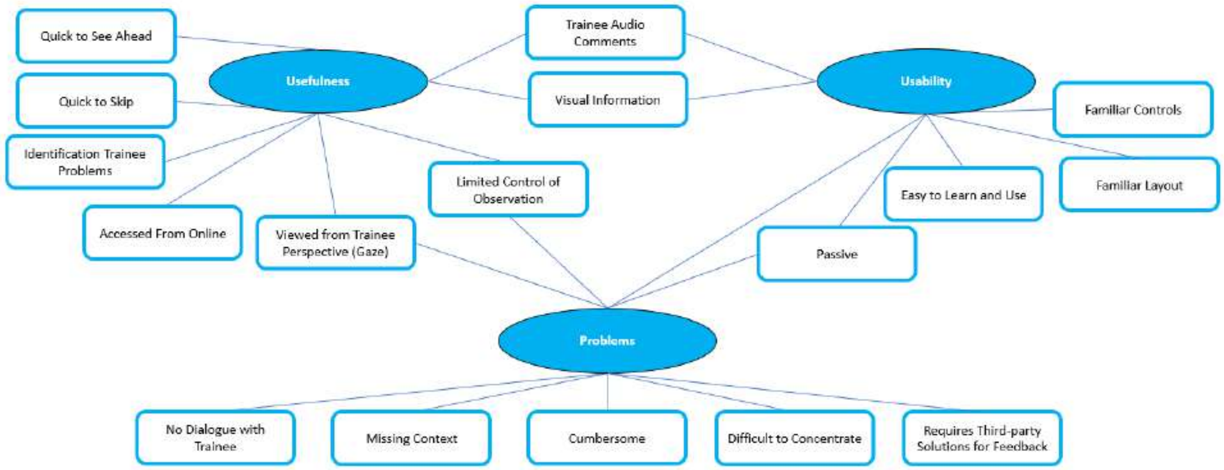


Figure E.3: Initial Video Assessment Thematic Map.

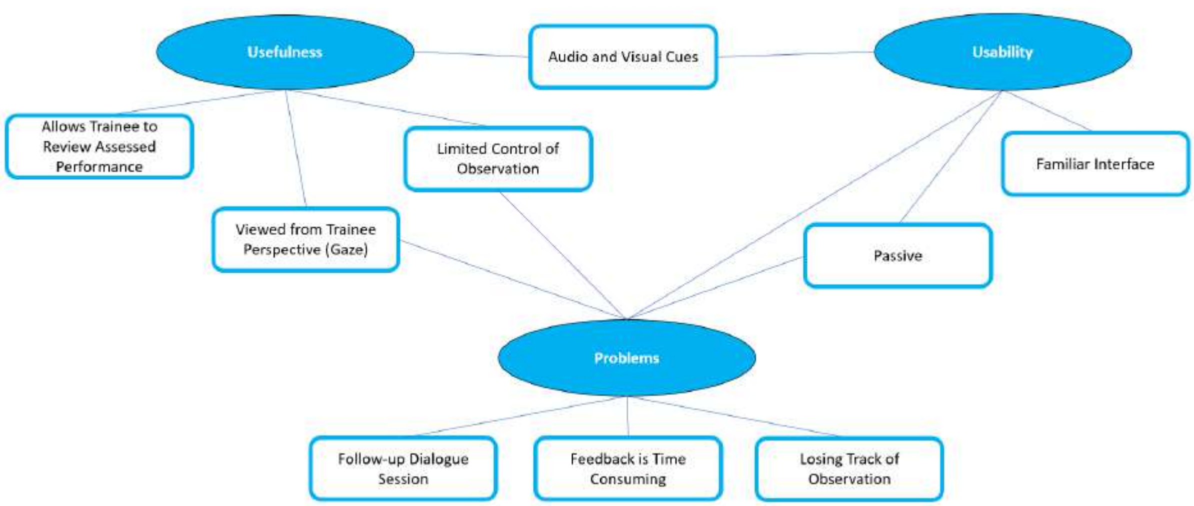


Figure E.4: Final Video Assessment Thematic Map.

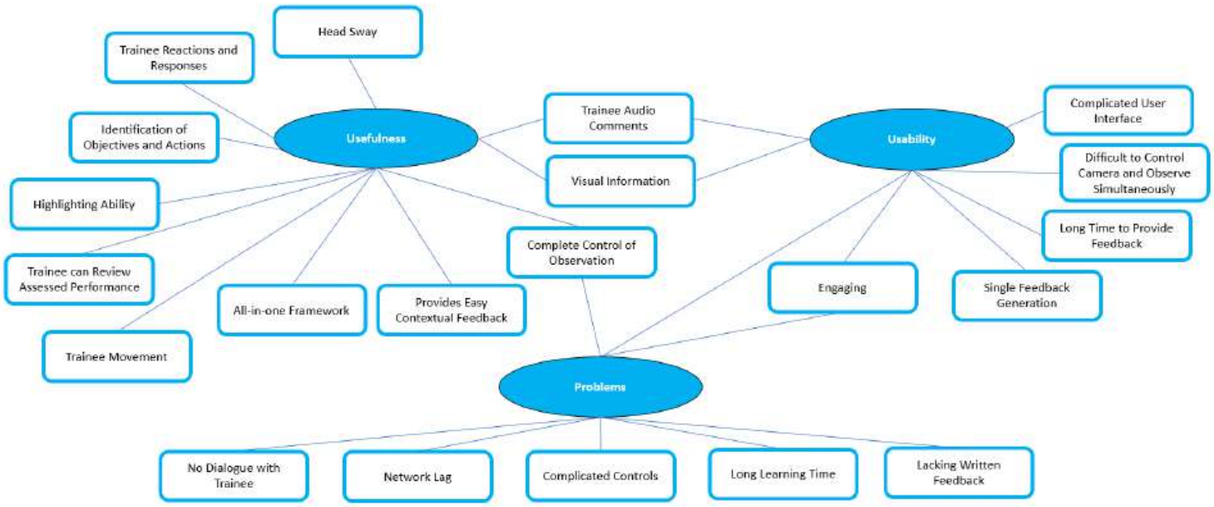


Figure E.5: Initial Framework Thematic Map

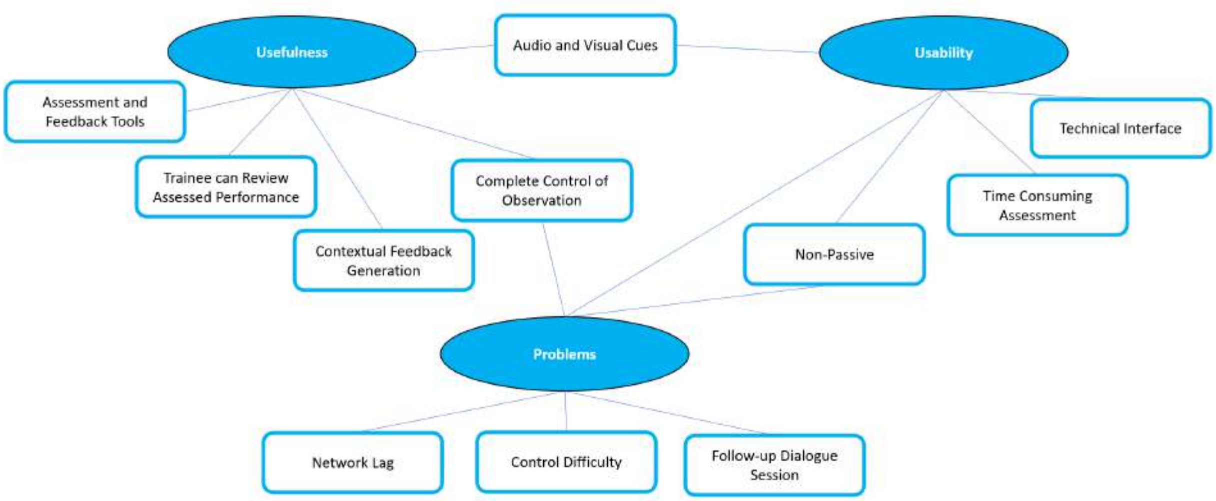


Figure E.6: Final Framework Thematic Map.

Appendix F

Previously Published Works

Virtual Observation of Virtual Reality Simulations

Scott Ronald Howie
University of the West of Scotland
Paisley, United Kingdom, UK
Scott.Howie@uws.ac.uk

Marco Gilardi
University of the West of Scotland
Paisley, United Kingdom, UK
Marco.Gilardi@uws.ac.uk

ABSTRACT

Unlike conventional desktop simulations which have constrained interaction, immersive Virtual Reality (VR) allows users to freely move and interact with objects. In this paper we discuss a work-in-progress system that 'virtually' records participants movement and actions within a simulation. This system recovers and rebuilds recorded data on request, accurately replaying individual participants motions and actions in the simulation. Observers can review this reconstruction using an unrestricted virtual camera and if necessary, observe changes from recorded input devices. Reconstruction of each participants skeleton structure was created using tracked input devices. We conclude that our system offers detailed recreation of high-level knowledge and visual information of participant actions during simulations.

KEYWORDS

Virtual Reality, VR, Observing, Replaying, Reviewing, Simulations, Virtual Observation

---

Woodstock '18, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-9999-9/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

INTRODUCTION

Reviewing VR simulations allows insight into participant performance [5] and determine the cause and effect relationships from participant actions [5]. Unlike desktop computers, VR equipment has six-degrees of freedom [8]. Because of VR's unrestricted movement, usability issues are common for inexperienced users [8]. To observe participants using VR an observer would need to be present to take notes or record the participant. This type of observation is difficult, requires large data storage, can impact the simulation performance and offers limited insight of the participants experience. Experiencing the exact perspective of participants induces motion sickness on the observer [7]. Observations can be conducted from mirrored perspectives of VR participants on 2D-screens. But these recordings can not guarantee knowledge of the state and location input devices when they are not visible limiting studies outside laboratory conditions. Video recording participants offers a visual clarity of their movement in VR, but are fixed in position and require to be paired and synced with other data to be of any use.

This paper details an initial validation phase on our work-in-progress into virtually observing participants in VR simulations. The term 'Virtual Observation' (VO) is used as we are not observing the participant themself, instead we observe the input movement and actions conducted by the user during the simulation. Collected input and movement data is used to re-create their exact experience in the simulation with synced actions and approximate movement. In the following sections we discuss our 'Virtual Observation' system and review the re-created simulations.

RELATED WORK

Research interest to review VR simulations is not new. Goldberg, Knerr and Grosse [4] used early implementations of VR technology to replay and review army training simulations in an effort to measure participants performance. More recently Lopez et al. [7] used a technique that creates and stores animation of movement and object manipulations conducted by the participant. Lopez et al.'s [7] system is restricted at re-creating the motions of the simulation, lacking high-level knowledge (user input or device information) of functionary events or actions. Without high-level knowledge it is difficult to detect usability issues in an interaction system.

Jung et al. [6] also developed a system that aims at creating re-usable animations for animating 3D character models. Jung et al.'s [6] recorded data posture combines the motion of the participant from tracking devices and pre-determined input states for grasping an identified object. Like Lopez et al. [7], Jung et al.'s [6] pre-defined input and interaction objects limit the high-level information that can be gathered from observing the animations.

Both Lopez et al. [7] and Jung et al.'s [6] methods are limited and for in-depth analysis of participants actions in a simulation high-level data (input and states) is necessary.

Figure 1: Rebuilding of simulation from recorded data. Input states are derived from this data and determine how the participant interacted with the virtual environment. Input states are overlaid from the Unity UI for visual clarity.

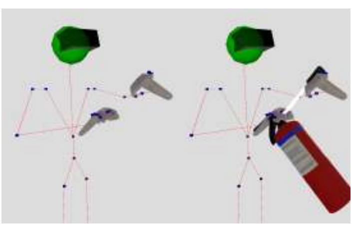


Figure 2: Screenshot of a participant in a default standing pose putting out a fire in the rebuilt simulation (see Figure 1 for alternative angle). (left) Participant skeleton pose. (right) Participant skeleton pose with held items.

THE VIRTUAL OBSERVATION SYSTEM

'Virtual Observation' (VO) combines unrestricted observation of participants [3] with a form of action capture [6, 7] allowing participants to be observed from any perspective. Rather than storing information of an entire environment, VO records participants movement and input changes. Unlike [6] and [7], recording does not include items or objects that are altered by the participant input, as they can be re-obtained by feeding the recorded inputs to the VR simulation as if the motions or actions were sent from input devices. VO allows individual simulations to be replayed, giving an insight on how exactly the participant acted during the simulation.

However, any object that has unpredictable movement, such as a non-player controlled character or sensitive physics affected objects, may not be the same during the re-built simulation. To address this, any unpredictable object can be tracked separately.

Positional and rotational data of the HMD and controllers is recorded at a fixed frame rate interval of 100 milliseconds. When a modified input state is detected, an input action key is recorded, which stores the motion of the controllers and the VR manager with the current input modifications. During the simulation VR configuration data is also stored to allow for tracking play-space and headset information to be analysed.

Rebuilt simulations use participant data-sets to replicate the movement and actions of the input devices, see Figure 1. Participant movement is smoothed between action and movement frames capture points to animate the movement of the controllers and HMD. During instances of input changes, the position and rotation of the controllers is forcibly set to ensure that interaction is correctly mapped at the exact position and orientation recorded and is not affected by any delays or gaps in the animation smoothing process.

Using Unity's Inverse Kinematic (IK) system, an estimated skeleton posture of the participant is created using the rebuilt tracking points from the VR equipment, see Figure 2, these tracking points map the head and hand locations of the participant to an avatar for real-time animation.

To observe participants, the observer controls a 3D virtual camera to explore the virtual environment. This allows to observe from any view the participants re-constructed actions and device motions during the replay. To reproduce a participant simulation the rendering and update of the Unity game engine had to be checked to prevent positional data from being set or timed incorrectly from the original data time-stamps. This required to modify the fixed update cycle of the Unity engine which we set to update every 0.01 seconds. The simulation should be assessed to ensure that frame intervals are acceptable for the needs of the simulation recording.

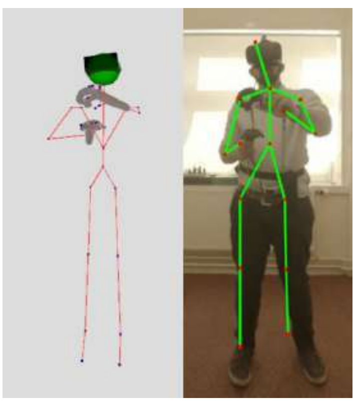


Figure 3: Side by side comparison of an estimated skeleton posture obtained from our system (left) and a participants posture from GoPro video footage with OpenCV's pose estimation (right).

METHODOLOGY

To evaluate our 'Virtual Observation' system we created a fire training simulation that aims to teach participants safety protocols. After obtaining ethical approval, we conducted a study with 12 participants in our lab controlled environment using an HTC Vive with a wireless TPCast attachment. Two GoPro Hero 4 cameras were setup at opposite areas of the controlled VR play-space area to video record a complete view of the tracked VR play-space and participant. Desktop screen-capture of participants perspective during the simulation was recorded at pre-determined interaction sections. In the training simulation participants were scored summatively on their performance from a grading system that assessed objectives completion.

Data was recorded anonymously using ID codes. Tracked motion and input data was instantly stored online making data available for analysis if a participant left or stopped the study early, offering a potential insight on what they did prior to exiting the application.

A training stage familiarised participants with controls and the necessary safety protocols. After successfully completing all training tasks, participants performed a search task in a different virtual environment aiming at locating and extinguishing fires.

To rebuild the simulation, participants data were loaded using the unique ID code, allowing to replicate participants use of the input devices and to rebuild the simulation as experienced by the participant. To assess and validate our virtual observational software, five frames were identified for each participant (see Table 1). Using Arbib's [1] action equation, we assessed simulation rebuilding accuracy by comparing similarities between the actual participant recordings and rebuilt recording during highlighted action frames (see Figure 4).

Assessment of the skeleton posture was conducted by comparing screen-shots from when the actions were started as indicated by the input state. GoPro footage was synced to match simulation time with screen-shots taken and processed using OpenCV's PoseEstimation with default settings to estimate each participants skeleton posture (see Figure 3), which was used as the valid structural target. These OpenCV estimations were then compared to same frame screen-shots from our estimation IK software. Joints were classified valid (1) if their relative positioning from connected joints was structurally similar to the OpenCV joint for position and bend. Miss-aligned joints on the Unity IK structure were marked as invalid (0). Percentages were calculated by adding all valid joints for each frame and dividing them by the total number of joints, giving the overall percentage for each frame joint. Any OpenCV skeleton posture that was visibly miss-calculated was removed (12 images were removed) from the analysis. Table 1 gives a scoring for the evaluation of each frame assessed.

FINDINGS

The ability to observe approximate poses and skeleton postures without restrictions offered insight into the motions of participants in the training simulation allowing to discover the how and potentially why a user failed to complete an objective. We were able to identify numerous cases of participants failing to interact with objects and/or using the wrong input command/button. We were able to observe that fire extinguishers were frequently mishandled, likely produced by 'magical interaction' [2] which removes the senses of weight and cumbersomeness from real-life equipment. Participants also frequently attempted to interact with objects in incorrect positions.

From a developer standpoint, the ability to view and replay the exact input, actions and movement of participants offers the chance to directly observe the process that lead to bugs or errors. In our case, we noticed that if a participant activated the fire extinguisher but then dropped it, the fire extinguisher would continue to be active spraying foam.

We observed numerous cases of people failing to follow instructions and protocols for safety procedures. Going by our pre-defined scoring system alone, which looked at goal completion, we would have incorrectly assumed all participants followed safety protocols as instructed. Half of our participants would have failed the real-life counterpart training for improper handling of safety equipment or dangerous movement.

Accuracy of the skeleton posture, see Table 1, varied by stance. Head and hands were tracked accurately with data provided from the VR equipment. The posture of the estimated skeleton derived from the position and rotation of the headset which all other joint estimations relied upon. Due to the lack of torso tracking the IK system occasionally miss-calculated the locations of the arm joints.

Although the skeleton reconstruction is not perfect yet, the strength of this work-in-progress is its ability to reconstruct the actions of the participant during the simulation. A demonstration of the reconstruction can be viewed here: youtu.be/NLGKC60qCGY, which shows a comparison between the rebuild simulation camera and the screen-capture recording, Figure 4 shows a frame of such comparison.

Developers and researchers can use VO to analyse feasibility and usability of systems. Examiners of training simulations can use the data to check and assess how participants have completed training procedures and verify they are correctly handling equipment/protocols. VO allows future comparison between participant performance or to assess if a bug present in the recording has been fixed. VO recordings can be easily stored and accessed when required. On average, each participant submitted a data-set of around 80MB for the simulation on reques which took ten to fifteen minutes to complete. Data-set vary depending on participants actions (input changes) and length in the simulation. In all cases, these data-sets were by large smaller than GoPro video recordings which were around 5.8GB for each participant (single camera) and screen-recording of around 500MB for 30 seconds.

Table 1: Skeleton evaluation in percentage from comparing of OpenCV's reconstructions from GoPro frames and estimated pose from VO. (Frames: 1. Opening garage door. 2. Opening house door. 3. Picking up fire extinguisher. 4. Extinguishing fire. 5. Pressing review simulation button)

Frame	1	2	3	4	5
Torso	88%	96%	94%	100%	92%
L-Arm	94%	94%	90%	100%	94%
R-Arm	96%	96%	96%	100%	98%
L-Knee	94%	96%	94%	98%	96%
R-Knee	92%	96%	96%	98%	96%

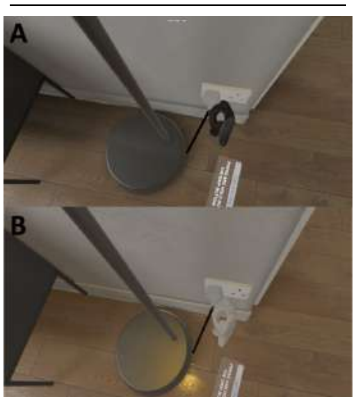


Figure 4: (top) Screenshot capture from screen-recording of participant. (bottom) Screenshot of the HMD FoV of the rebuilt simulation from the same participant.

CONCLUSION

This paper details our work-in-progress VO software. VO records data from VR simulations and reconstructs the experience using users actions and motions. Our findings indicate that rebuild simulations were identical to the original simulation (see Figure 4). We conclude that these recordings are on par with [7], but offer greater insight into high-level technical details of the simulation, offering context from cause and effect relationships of participants actions [5].

Using only two controllers and a VR headset as the tracking devices, the reconstructed skeleton structure of participants varied in quality depending on the posture of the participant. Lack of torso tracking often limited accurate estimation of the body orientation and stance.

Lack of finger tracking on the controller limits the observation of controllers usability as non-registered input on the controllers can not be tracked.

We conclude that virtual observation of reconstructed simulations is a highly effective and versatile tool for researchers and developers, with reconstructed participant actions identical to real performance. In the future an 'in the wild' study will be conducted to assess the viability of virtually observing participants outside of laboratory conditions. Further on this research will be expanded to determine if and how such a system could be incorporated directly into a evaluation system that provide instant feedback [5]. Additional tracking devices and features (eye-tracking, torso, audio) will be included to improve data acquisition.

REFERENCES

[1] Michael A Arbib. 2002. fO The Mirror System, Imitation, and the Evolution of Language. *Imitation in animals and artifacts* 229 (2002).

[2] Doug A Bowman, Ryan P McMahan, and Eric D Ragan. 2012. Questioning naturalism in 3D user interfaces. *Commun. ACM* 55, 9 (2012), 78–88.

[3] Joel Carranza, Christian Theobalt, Marcus A Magnor, and Hans-Peter Seidel. 2003. Free-viewpoint video of human actors. In *ACM transactions on graphics (TOG)*, Vol. 22. ACM, 569–577.

[4] Stephen L Goldberg, Bruce W Knerr, and James Grosse. 2003. *Training dismounted combatants in virtual environments*. Technical Report. Army Research Development and Engineering Command Orlando FL Simulation âĂę.

[5] Samer Hanoun and Saeid Nahavandi. 2018. Current and future methodologies of after action review in simulation-based training. In *Systems Conference (SysCon), 2018 Annual IEEE International*. IEEE, 1–6.

[6] Bernhard Jung, Heni Ben Amor, Guido Heumer, and Matthias Weber. 2006. From motion capture to action capture: A review of imitation learning techniques and their application to VR-based character animation. In *Proceedings of the ACM symposium on Virtual reality software and technology*. ACM, 145–154.

[7] Thomas Lopez, Olivier Dumas, Fabien Danieau, Bertrand Leroy, Nicolas Mollet, and Jean-François Vial. 2017. A playback tool for reviewing VR experiences. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. ACM, 83.

[8] Jolanda G Tromp, Anthony Steed, and John R Wilson. 2003. Systematic usability evaluation and design issues for collaborative virtual environments. *Presence: Teleoperators & Virtual Environments* 12, 3 (2003), 241–267.

Virtual Reality
https://doi.org/10.1007/s10055-020-00463-5

ORIGINAL ARTICLE

Virtual Observations: a software tool for contextual observation and assessment of user's actions in virtual reality

Scott Howie[1] · Marco Gilardi[1]

Received: 29 November 2019 / Accepted: 27 July 2020

Abstract
In this paper, we present 'Virtual Observation' (VO) a software tool for contextual observation and assessment of user's directly from within the virtual reality (VR) simulation framework. Unlike other recording systems, the VO system described in this paper focuses on recording and reconstructing VR user's positional, rotational and input data to recreate the same experience the user had with a VR simulation. Different from animation-based approaches, VO records user inputs and reconstructs the simulation from them and the user positional data. Moreover, the system allows the broadcast of this information to a remote machine enabling remote live observation of the simulation. Datasets recorded by the system can be shared by exporting them as XML files or, optionally, into a standalone online application, such as browser WebGL, allowing researchers, developers and educators to share and review a VR user simulation through a free-moving camera using a web browser. In this paper, the consistency of the data generated from the software by the client, server and reconstructed datasets acquired during real-time live observations was evaluated. We conclude that this Virtual Observation software offers detailed reconstruction of low-level information and visual information of user actions during simulations for both live and offline observations. We envision that our system will be of benefit for researchers, developers and educators that work with VR applications.

Keywords Virtual reality · Virtual Observation · Observing · Replaying · Reviewing · Simulations

1 Introduction

Direct observation of users in their context can offer insight on design challenges (Lazar et al. 2017) and circumvent the issue often encountered of users describing inaccurately what they did due to a lack of awareness or understanding of the task or system under study (Blomberg et al. 2007). The ability to replay and review user's sessions is an important tool for assessment (Lazar et al. 2017). However, when it comes to studying users in virtual reality (VR), performing contextual studies becomes challenging, as the user and the observers are positioned in two different contexts: the observer is in the real world, whilst the user is in the VR context. Observing users from inside the same VR simulation allows insight into their performance (Hanoun and Nahavandi 2018) and helps in determining the cause and effect relationships from user actions (Hanoun and Nahavandi 2018). Unlike desktop computers, VR equipment has six degrees of freedom (Tromp et al. 2003), and because of VR's unrestricted movement, usability issues are common for inexperienced users (Tromp et al. 2003).

Currently, to observe users using a VR simulation an observer needs to be physically present to observe the user's body movement in the real world, take notes on what the user does or video record and screen capture the user. Although rigorous, this type of observation is difficult, requires large data storage for the videos, can impact the simulation performance and offers limited insight of the users experience and interaction. The ideal solution would be to experience the exact perspective of users, but this approach has the drawback that it can induce motion sickness on the observer (Lopez et al. 2017). Observations can be conducted from mirrored perspectives of VR users on 2D screens, but these recordings cannot guarantee knowledge of the state and location of the input devices when the user

✉ Scott Howie
Scott.Howie@uws.ac.uk

Marco Gilardi
Marco.Gilardi@uws.ac.uk

1 University of the West of Scotland, Paisley, UK

is not looking at them, limiting studies inside and outside laboratory conditions. Video recording users offer visual clarity of their movement in VR, but the point of view of the recording camera is fixed in position and user actions can be obscured from the camera point of view. Moreover, this approach requires the video data be paired and synced with simulation data, such as 2D-screen recordings of the VR perspective before it can be analysed.

This paper builds upon Howie and Gilardi's 'Virtual Observation' system described in Howie and Gilardi (2019). Howie and Gilardi (2019) system records and reconstructs virtual actions of users in a simulation, allowing researchers and developers to virtually observe users in a VR simulation. The term 'Virtual Observation' (VO) is used as observers are not observing the user directly, but instead they observe the input, movements and actions conducted by the user within the simulation. Collected input and movement data is used to recreate the user experience in the simulation with synced actions and approximate movement. In Howie and Gilardi (2019), VO was validated in a detailed simulation environment for fire safety which tasked the participant to correctly identify and extinguish fires in a VR training simulation (Howie and Gilardi 2019) (Fig. 1). VO was used for reconstructing users movement and actions for review *after-action*, meaning the observation was conducted after the user had completed the simulation (Howie and Gilardi 2019). The VO system in Howie and Gilardi (2019) can reconstruct VR simulations for a wide range of use cases (Wobbrock and Kientz 2016).

Building upon the initial version of VO (Howie and Gilardi 2019), we expanded the system functionalities to

Fig. 1 Screenshot captured from video comparison showing real-time video capture of a simulation and rebuilt capture using Virtual Observation. The full video can be viewed here: https://www.youtube.com/watch?v=2YY-d7QMUVI . The Virtual Observation system can be included in any SteamVR compatible Unity project by adding the VO component to the default VR camera rig set-up and linking the input states of the developers/researchers interaction system (so the actions can be recorded and later replicated). Recorded data can be stored in XML or JSON file formats and supports all VR headsets and tracked controllers that are compatible with SteamVR and the Unity game engine

allow for remote live observation of users via WiFi from an external location in real time using Unity's default networking architecture, added configuration for full body tracking, optional hand tracking and an 'Full Simulation Capture' recording mode that can record and reconstruct external tracked object classes and variables. Moreover, controls such as play/pause and rewind/fast-forward were added to allow observers control over the reconstructed simulation.

Although there are other popular platforms that facilitate the development of VR applications, such as the Unreal Engine, we chose to develop the system using Unity. This choice was made for convenience; Unity is one of the popular and widely used platforms for developing VR and the platform with which authors are most familiar. However, the principles underlying VO, i.e. recording user input and positional tracking data to reconstruct the simulation, can be transferred to other engines, and Unity should be seen throughout the paper as the mean that we chose to prove that a system such as VO can be developed.

This paper contributes to the field of virtual reality by presenting a new approach to recording VR simulations for training scenarios. The approach is implemented as a proof of concept of a versatile tool that will allow researchers, developers and educators that use VR, to observe, evaluate and share user actions and interactions within a VR simulation, either live or by reconstructing it via recorded data. Moreover, we demonstrate that:

1. data generated by VO can be reliably streamed to a remote machine for real-time live observation of a user VR session,
2. streamed VO data can be recorded in the observer's remote machine and that simulations reconstructed from this data are identical to the VO data recorded on the user machine during the user simulation, showing VO reliability in recording VR user sessions,
3. VO data can be used to share the VR user session after it has ended on both VR and non-VR platforms, such as a WebGL visualiser.

2 Definitions

The following terms will be used in this paper:

– *Motion key-frame* values are the position, rotation and scale of the tracked devices in the virtual simulation space. Motion key-frames are recorded through-out the duration of the simulation in sequential order, later being used to reproduce the movement of the VR user during reconstruction. We refer to motion key-frame data as *high-level information* which is acquired from real-time values from the Unity transform component during a simulation.

- *Action key-frame* are frames in which input from one of the controller devices has been modified from a previous state, for example, a button on a tracked VR input device that changes state from idle to being pressed down would be recorded as an action key-frame, since a variable that derived from a VR input devices had been altered. We consider action key-frame data as *low-level information* since the data acquired is in a *raw data* format of values that correlate as one's, zero's or customised values, with input values acquired or derived from the VR controller button state changes. During reconstruction, the action key-frames are reconstructed to replicate the experience of a VR controller device being modified.
- The terms *Rewind* and *Fast-forward* are used to describe the functionality in the VO system that allow users to jump to a specific time frame, before or ahead of the current time position, and see the simulation from that point onward. The terms are used for a lack of better words to describe the functionality and should not be confused with being able to see the simulation replayed backwards or at accelerated speed.

3 Related work

The problem of finding a way to review VR simulations is not new, and systems similar to VO have been previously proposed in the literature. One of the earliest examples is Goldberg et al. (2003), which used early implementations of VR technology to replay and review army training simulations in an effort to measure users performance. Goldberg et al. (2003) after-action review system used event data collection processes to capture the events that took place during a simulation. Whilst this offers a guarantee that an event can be triggered repeatedly during the after-action review, the input of the VR user is not recorded. Therefore, any usability issues experienced by the VR user cannot be replicated and therefore prevent usability analyses to be conducted.

Greenhalgh et al. (2002) developed a technique of temporal links which enabled the recording of data in the MASSIVE-3 system (Greenhalgh et al. 2000), with the objective to link prior recordings to real-time VR environments. Greenhalgh et al. (2002) system operates by recording all changes made to the virtual environment at run time. The focus of Greenhalgh et al. (2002) work was integrating VR data for media use in film and television and incorporation of previous VR experiences into a real-time VR environment to create new content or review experiences. Using temporal links, Greenhalgh et al. (2002) method allowed recorded virtual environment can be replayed and embedded into a live virtual environment for purposes of extending the live virtual environments content or narrative structure. Although not stated, the implementation of temporal links (Greenhalgh et al. 2002) appears to be designed for desktop VR and not head-mounted display VR. Likewise, the description of the temporal links system (Greenhalgh et al. 2002) hints that the recording and reconstruction of the temporal links use animation-like recording of the virtual environment rather than raw data interpretation of the VR users as used in VO, which enables usability testing and repeatable interactions for development and bug fixing. In a use-case example for reviewing virtual experiences (Greenhalgh et al. 2002), the authors describe the ability to replay the virtual environment from any perspective but do not discuss any ability for interaction of user input or variable monitoring of recorded objects. This lack of clarity suggests the temporal links system is animation based which suffers from the same limitations as other animation reconstruction systems (Lopez et al. 2017), such as the lack of raw user-input data, which limits the understanding of usability issues from input devices.

Similar to Greenhalgh et al. (2002), Von Spiczak et al. (2007) expanded upon the OpenTracker framework (Reitmayr and Schmalstieg 2005) to create a multi-modal event processing system. A feature of Von Spiczak et al. (2007) approach is the ability to capture position, orientation and interactions of equipment from the multi-modal event data structure. This data structure can be used to replay, review and document VR interactions in a virtual environment in a serialised order. These recordings captured positions and orientations of the tracked objects along with other interactable object information within the virtual environment. When replayed, these simulations would reconstruct the event with a time-delay between the captured data points. Von Spiczak et al. (2007) show that their system is viable, but only discuss briefly the reconstruction of interactions in their paper, using the data reconstruction example as a use case for their multi-modal event processing system designed for the OpenTracker framework (Von Spiczak et al. 2007; Reitmayr and Schmalstieg 2005). Von Spiczak et al. (2007) make no mention of rewind or pause ability, observational potential, live broadcasting of data, portability of their system and functionality of the system beyond the ability to replay and log the VR interactions in real time. Both Greenhalgh et al. (2002) and Von Spiczak et al. (2007) implementations are designed as extensions for specific systems and are incompatible with the Unity platform, which relies on game engines. In comparison with VO, the discussed implementations (Greenhalgh et al. 2002; Von Spiczak et al. 2007) share some methodological similarities in approach for capturing and reconstructing actions; however, Von Spiczak et al.'s (2007) study lacks important functionalities for allowing effective observations of VR users, such as: control the perspective, time and playback of the reconstruction, live observation, automatic data serialisation, recording and reconstructing objects and variable data beyond the VR interaction devices, lightweight portability of data and multi-user data capture.

More recently, Lopez et al. (2017) used a technique that creates and stores animation of movement and object manipulations conducted by the user. Although effective, Lopez et al. (2017) system is restricted at re-creating the motions of the simulation and does not store low-level information about the interaction. Because Lopez et al. (2017) system stores all data into animation files, the data can only be accessed using the Unity game engines animation system, preventing statistical data to be collected for state or positional analysis in external programs.

A system similar to VO was created by Jung et al. (2006). Although Jung et al. (2006) system is capable of creating reusable animations for animating 3D character models, it only records predetermined input states for grasping an identified object. Unlike our VO system, which aims to capture all states from VR input devices, such as the state of input buttons, including raw values of analog inputs. Jung et al. (2006) system focuses on storing the grasp events so they can later be imitated for manipulating virtual scenes with different grasping types. Like Lopez et al. (2017), Jung et al. (2006) pre-defined input and interaction objects limit the high-level information that can be gathered from observing the animations, with neither systems aiming to capture all states of the VR input devices. Both Lopez et al.'s (2017) and Jung et al.'s (2006) methods are limited as for in-depth analysis of user's actions in a simulation knowledge of high-level information (input and devices states) is necessary.

Alternative 'virtual' means of observation can be used to monitor participants in VR remotely through the use of cameras and remote screen sharing (Lazar et al. 2017), but these options cannot guarantee knowledge of the state or location of tracked devices. Even with these forms of observation available, for remote studies conducted 'in the wild', setting up software or hardware for observing remote users could be a challenge (Petrie et al. 2006). These observation methods also require the participant to have access to suitable recording hardware and accompanying software for remote observations to be possible.

Video and audio recordings are a common way for studies in laboratory and 'in the wild' to document experiments and procedures that can be then analysed qualitatively (FitzGerald 2012). These recordings aim to make it possible to analyse the events in the experiment after the experiment has been conducted using a format that is easy to share between collaborators and for demonstration purposes (FitzGerald 2012). Our VO system adopts (FitzGerald 2012) approach and innovates the medium to a digital data format designed for recording and reconstructing VR simulations that can be shared independently of a development system for browser-based collaboration. Parallel to transcription work that is often applied to audio and video recordings (FitzGerald 2012), VO allows for software detection of user input and movement, that in the future, could be paired with AI for analysis of participant actions.

Different from the systems discussed, VO focuses on the low-level acquisition and reconstruction of data in an easily controllable dataset that once implemented into an application automatically and non-invasively records the actions of a participant in a VR simulation, without any modification by the developer or end-user. In VO, the recording of user inputs depends exclusively on the input controllers SDK (OpenVR, Oculus SDK, Leap Motion or other third-party SDK) chosen for the VR simulation, allowing different types of controllers to be supported, as shown in Fig. 3 where hand tracking is used in place of the VIVE controllers, and new controllers can be easily added in the future to adapt to the fast evolving VR input modalities. Finally, rather than replaying a VR simulation as an animation (Lopez et al. 2017; Greenhalgh et al. 2002), VO reconstructs the experience using the low-level data of the VR users movement and input. Moreover, the system is designed to be lightweight, portable and user-friendly and it is developed using a modern game engine, namely Unity.

4 The Virtual Observation system functionalities and design

The VO system presented in this paper is a combination of unrestricted observation of users (Carranza et al. 2003) with an improved form of action capture (Lopez et al. 2017; Jung et al. 2006) and allows users to be observed from any perspective. Rather than storing information of the entire environment, the VO system records users movement and changes in low-level information data captures. The low-level information captured is configurable and as default supports all OpenVR hardware input devices, but can easily be extended to suit other input devices; for instance, Fig. 3 shows the integration of Leap Motion. A callback pattern allows VO to listen and dynamically record extended input, such as additional API input information. In contrast to Jung et al. (2006) and Lopez et al. (2017), which record the motion of the interacted virtual objects directly, the major contribution of the VO system is that only users inputs, device states and tracking information is recorded by the system. The VO system captures and records an unconstrained number of tracked objects and devices in addition to three tracked main VR devices: head-mounted display (HMD) and two controllers. By adding additional tracked objects (Fig. 2),

non-player controlled objects such as AI characters or non-deterministic physics-affected objects can be reliably

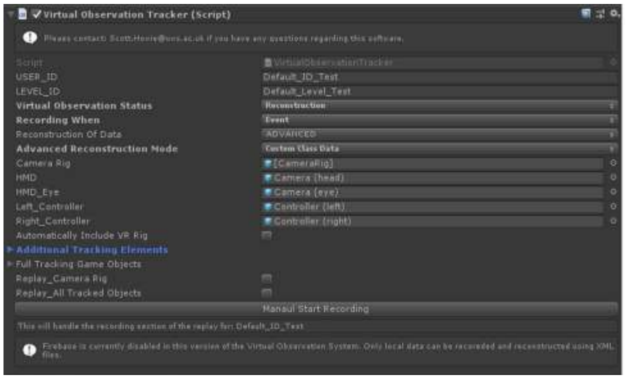


Fig. 2 The VO system is built for the Unity game engine. The component can be attached to any VR player object with options for HMD, HMD Eye and controllers the default options for tracking of movement. Additional tracking objects offer extension to the prerequisite tracking objects. The variable 'Full tracking Game Objects' is used for tracking a hierarchy of objects, this will track the parent object and all subsequent child objects

Fig. 3 Screenshot captured from a reconstructed simulation that used a leap motion tracking device to capture the movement of a user's hands. This demonstrates VO's ability to extend recording and reconstruction to a wide range of interaction devices. A video demonstration of the VO system can be viewed here: https://www.youtube.com/watch?v=IX1qjq1R134

tracked, ensuring that the reconstructed simulation is consistent with the simulation 'as-experienced' by the user. Unrestricted tracked objects and devices also allow for novel or unconventional objects to be tracked, expanding the scope of the tracking hardware to include additional interaction systems that rely on eye-tracking or hand-tracking (Fig. 3).

The proof of concept for the VO system was implemented as a plug-in for the Unity game engine and can operate in two modalities, 'Interaction Capture' and 'Full Simulation Capture'. The *Interaction Capture* mode records and reconstructs VR users position, rotation and actions within a simulation, isolating data captured to modifications made by a single participant. However, as the 'interaction capture' recording technique does not capture non-deterministic objects, an advanced method which can be toggled to automatically capture all gameobjects that contain non-deterministic factors, such as AI or physics impacted objects, was added to the system. The *'Full Simulation Capture' mode* extends upon the 'Interaction Capture' mode by including serialisable and custom-assigned variables of additional tracked objects, which are optional and can be used for the reconstruction of AI driven or physics gameobjects that are non-deterministic.

4.1 Interaction capture

To record a VR simulation in VO, positional and rotational data of tracked devices are motion key-frames that are recorded at a fixed frame rate interval of 10 milliseconds, which is the lowest value of the record and reconstruction functions being called during a VR simulation (see Howie and Gilardi 2019). Input action key-frame when a modified input state is detected is also recorded, along with all non-deterministic game data, such as modified objects with physics properties. The recording process stores the current transform of all tracked devices in the VR manager with the current input modifications. VR tracking configuration data, such the width and depth of the physical assigned VR area, is also stored to allow for tracking play-space and headset system information to be analysed. Tracking configuration data allows to contextualise the tracked movement of users relative to their configured tracking space in the real world during 'in the wild' studies.

4.2 Full simulation capture

The prototype of the VO system in Howie and Gilardi (2019) was originally intended for observation of pre-defined training simulations, which only required the knowledge of the VR user to operate (Howie and Gilardi 2019). In this paper, we present an extended version of VO that includes a 'Full Simulation Capture' method of recording external actions of other non-player controlled objects. Because some simulations rely on the knowledge of external human interaction factors to regenerate the simulation or AI characters, additional objects and their attached components can also be recorded and reconstructed to be in sync with the VR user. This recording process can automatically serialise the additional tracked component data of serialisable classes and user-defined data from custom or protected Unity components (i.e. Rigidbody or Colliders). This data can be recorded in sync with the VR user recording motion key-frame. This 'Full Simulation Capture' mode is less user-friendly than the 'Interaction Capture' mode because as it requires the developer to specifically detail each component and variable that is required to be recorded and reconstructed during development. At present, 'Full Simulation Capture' mode uses customised classes to store the active status, tag and name of objects, accessible 'rigid body' and 'collider' variables along with fully serialisable custom classes of tracked objects.

4.3 Reconstruction system

In order to reconstruct the transform data (position and rotation), data-points are captured during the recording of the simulation at motion key-frames and action key-frames. Motion key-frame data points are captured every 10ms, which was found to be reliable for movement estimations, whilst action-frames are captured when an input state change is registered. During reconstruction, the system determines whether the reconstruction time matches a recorded action key-frame or motion key-frame (Fig. 4).

During an action key-frame, the input states of the tracked devices are reconstructed by rebuilding the data to a class structure readable by Unity and assigning them to the object that matches the tracker. In our case, this was a VR Input management class, which handled the input of the VR controllers that determined the action, values and states of our interaction system. When an action key-frame or motion key-frame event is called during the reconstruction process, all tracked objects are set to the position recorded during the recorded key-frame. The software modifies the tracked objects based on the current position of the tracked objects and next transforms location based on the next action key-frame or motion key-frame data point. This technique enables the motion of the participant between timed gaps in the data to be estimated from the previous and next action or motion key-frames. See Fig. 4 for a diagram of the reconstruction process.

To determine whether the simulation is recording or being reconstructed, conditional checks are used in the VO system along with individual information, thus determining whether a user has direct control over an object, i.e. the user is holding or interacting with the gameobject. These checks are primarily designed to let the system know that simulation data is expected to vary from recording data and prevents individual object recorded data to override the user-driven input actions. During reconstruction, this does not matter as the user-driven input actions are reconstructed before any other game data; therefore, any variable modification to held objects will remain consistent. As such, the checks are intended for aiding the debugging process of the simulation.

4.4 After-action review

To make it possible to review a simulation at a later date or time than the original simulation, an 'after-action review' (offline) process is implemented. Such system focuses on recording the data for reconstruction after the user has completed the simulation. Action events that are registered from a change in input state are recorded along with the current position of all tracked objects. Key-frames for action and motion are sequentially logged in order of time gap from the start of recording and stored immediately on a database or recorded locally to XML (once the simulation has finished). This data can be retrieved and used for a simulation to be reconstructed when desired by the developer or researcher.

To replicate the movement and actions of the input devices, rebuilt simulations use user datasets. User movement is smoothed between action key-frames when input is modified, and motion key-frames that capture the motion of the tracked devices. During instances of input changes, the position and rotation of the tracked devices is forcibly set to ensure that interaction is correctly mapped at the exact position and orientation recorded and is not affected by any delays or gaps in the animation smoothing process.

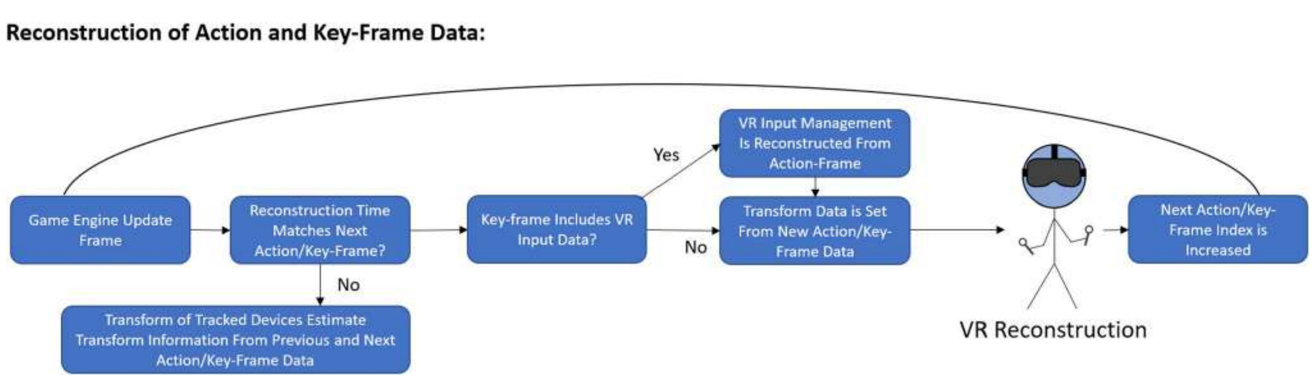


Fig. 4 The reconstruction process of the VO system, demonstrating how the reconstructed VR character transforms update based on the action or motion key-frame data. Transform data of tracked devices are recorded every 10ms as motion key-frames, whilst input management data and transform data of all tracked devices are recorded as action-frames when the a change of input state is detected

4.5 Live-action review

'Live-action review' is the process of virtually observing a user in a VR simulation in real time. This method is based on standard networking set-ups for multi-player games, in which a main user hosts the network-configured game acting as the server, with other players able connect to as clients, allowing data to be transferred continuously between client and server. The data structure and process of recording and reconstructing the data are the same as after-action review, but instead of storing it for later use, the data is streamed in real time from the client (user) to the server (observer). This process updates the simulation in real time by monitoring the actions and movement of the user from an external location. During live observation, actions and movement of users of the simulation who are clients on the server can be recorded locally, enabling data that is streamed to the server during the live observation to be stored using the same process as after-action review. Live-action review uses the Unity game engine networking suite to create the area for multiple clients or observers to be present within a simulation. Whilst only assessed with one client and one observer, in principle the software scales with Unity's multi-player functionality to support multiple members for both roles (simulation user or observer).

The VO system can be incorporated into existing infrastructures, requiring only one component class for data allocation. This approach allows the system to use Unity 2018's standard Networking Interface UNET for live transmission of data. Unity has since deprecated this Networking Interface, but the software will be compatible with Unity's replacement Networking architecture or any alternative that allows custom data transmission in real time. Live-action simulations use the networking data from the Unity network transform components to update the transforms of the tracked objects on the server (Fig. 5). These transforms have a send rate of 25 network updates per second, but locally saved movement recording of the tracked devices remains at 10 ms. When an input state change is registered on the client and sent to the server, all tracked objects on the server (including any other connected clients) are updated to their recorded position when the input state change was detected. This ensures that the transform of the objects during a received input action is consistent with their position on the client and is not affected by any lag or networking issues. Due to University firewall settings, we were only able to perform tests using a WiFi local area network (LAN); therefore, lag may be present when applied in different networks configurations depending on the connection of the server and client. Different network connections may cause potential issues over long distances or bandwidth limitations. However, we show in this paper that these issues are not caused by the VO system and are only dependant on the networking architecture used.

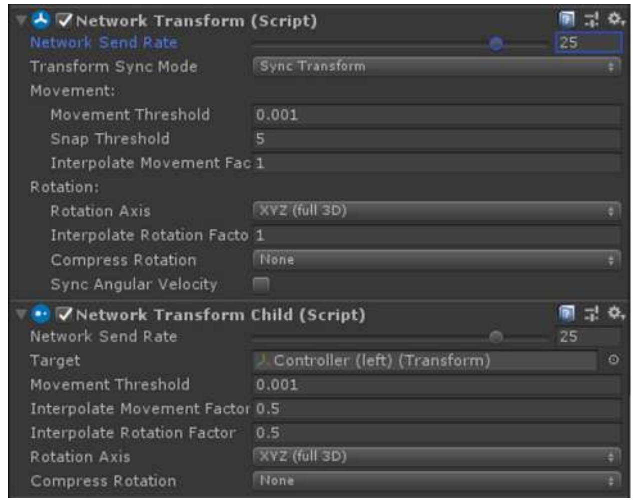


Fig. 5 Unity's networking configuration for sending packets of data between client and server in real time. Each player VR character is set as a local player authority on the network, with network transforms attached for the top hierarchical object, and child network transforms for all tracked child objects of the VR character. Data packets were sent for the tracked objects at a send rate network cycle of 25 updates per second

4.6 Reviewing observations

Using an inverse kinematic (IK) system, an estimated skeleton posture of the user can be generated using the rebuilt tracking points from the VR equipment (Fig. 8); these tracking points map the head, hand and other tracked VR locations of the user to an avatar for real time or after-action animation. To implement this, we used the FinalIK (Lang 2019) package for full-body tracking of the participants head, hands (controllers), torso and feet.

To observe users, the observer can either control a virtual 3D camera or use a separate HMD attached to the server computer (Fig. 6). Both forms of observation explore the reconstructed simulation in real time. To reproduce a user simulation, the rendering and update of the Unity game engine had to be mimicked to prevent positional data from being set or timed incorrectly from the original data timestamps. This required to modify the fixed update cycle of the Unity engine which we set to update every 10ms to match the recording interval.

Time-frame reproduction was improved with respect to Howie and Gilardi (2019) system, obtaining smoother transitions of action key-frame and motion key-frame capture points in after-action review reconstructions. One issue in Howie and Gilardi (2019) system was that a frame was skipped between the action and reproduction of the input

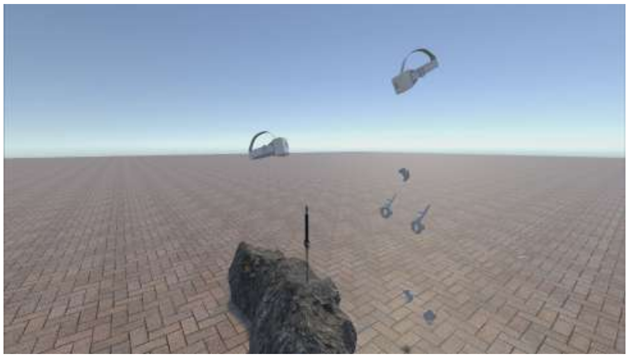


Fig. 6 Observers can observe users using a virtual 3D desktop camera or by equipping a VR HMD. Observations can take place live in real time with the user, or at a later stage for after-action observation. User datasets can be observed using the desktop system here: https://virtualobservationsystem.github.io/VirtualObservationSystem/

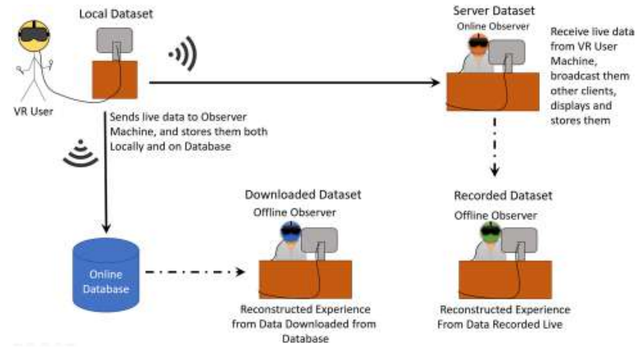


Fig. 7 The VO system has multiple methods of storing and retrieving data. For live observation, data is streamed to the server, updating the simulation in real time and storing the received information locally for offline observation later. For after-action review (offline observation), data is stored on an online database (Firebase) for retrieval by the server when necessary. A local XML copy can also be stored on the clients PC for recovery in local tests

which was caused by Unity's update pipeline. By using a modified order to the script execution pipeline, the skipped frame can be avoided with actions registered in the same action key-frame as the recorded input. To achieve this, the script execution order in Unity is changed so that the script that rebuilds the simulation is given priority over the input action scripts that determine the controller input actions; this ensures that the movement transforms and input state of the devices are exact during the time of an input state change. The execution of physics interactions for collisions is staggered during the reconstruction so the physics and events of objects replicate the expectations of the engine, with movement for physics collisions and triggers executed prior to input modifications.

The VO system is designed as 'drag and drop' component in Unity; this design choice was taken to simplify the integration of the system in VR simulations. The replication of movement can be achieved by adding the VO system to the top-hierarchy object of the player character and assigning the desired tracked objects (Fig. 2). Every VR player controlled character in a scene is individually recorded, independent of other users. This streamlines the process of storing and re-accessing data that is only applicable to a single-user performance and limits damage caused by bandwidth constraints or data corruption. For group simulations where an observed user interacts with other users in a single scenario, each individual dataset for all users can be loaded and reconstructed simultaneously, reproducing the movement and actions of all users within a single observation session. Recorded audio group communications can also be reproduced as VO records from the microphone input of each user if a microphone is attached to their VR headset and it is active. The audio can also be isolated for individual users, which is useful in scenarios where group communication makes it difficult to hear an individual user speaking.

Data can be acquired in several ways. For live observation of users, data streamed to the live observer can be recorded locally on the server. When data is received by the server during live observation, the reconstruction of the data is processed in real time and stored by replicating the data capture process that is being conducted on the client side. Alternatively, data can be acquired from the client by retrieving a locally stored XML file, avoiding the need for online database hosting. If a study is being conducted in a remote location, making the retrieval of data from the client computer impossible, an online database can be used to host the client data which can then later be retrieved by the observer when required. See Fig. 7 for a diagram of the data storage options.

4.7 Review functionalities

Review functionalities for observers allow to control the reconstructed simulation playback (play, pause, fast-forward and rewind) and observational position during and after-action review. These functionalities work independently of any external VR plug-ins or frameworks, allowing observations to be conducted on any Unity supported platform. In this paper, we show VO being used to observe a recorded VR simulation within a VR application, a desktop application, WebGL web browser hosted on a website and within the Unity game-engine itself and show the potential of VO being used for observational purposes in multiple fields.

Rewinding and Replaying a reconstructed simulation using the 'Full Simulation Capture' mode will result in repeated game events being re-simulated. During rewind instances, modified variable information will be restored back to previous values which will then be re-simulated when the reconstruction resumes in forward playmode.

Assuming both a fire extinguisher and fire were recorded along with the VR user, a reconstruction operating in rewind would reverse and the fire extinguisher capacity and the fire's health back to their previous values, along with all other recorded data recorded with the individual gameobjects. When resuming to play in forward mode, the simulation would reconstruct the actions of the user extinguishing the fire, since the variable information of the gameobjects was reset and the user-driven actions of using the fire extinguisher are consistent. This means in the fire training simulation demonstrated in Fig. 1, the fire could be extinguished and made re-active by the rewinding and re-simulating the experience.

Observing a training simulation within VR replicates real-world attributes of monitoring a trainee (see video in Fig. 9). Moreover, the option for observers to be invisible during the observation avoids users' feelings of discomfort caused by the physical presence of the observer. In VR, both the observer and VR user share the same context, both experiencing the presence and scale of the virtual environment. Unlike in the real world, observing in VR is non-intrusive, with the ability to monitor the training experience from any perspective without interfering with the VR training user's performance.

In desktop applications (Fig. 9), the observation is controlled using keyboard and mouse input and operates closer to the functionality of video playback, but with the functionality of altering the viewing perspective of the observation camera. This observational method has greater accessibility for observers since the application can be operated using standard computer hardware, with no additional hardware or programs beyond the simulation application.

Like the desktop implementation of VO, the WebGL browser functionality is an accessible and portable method of observing VR simulations (Fig. 6). WebGL applications can run on any web browser platform that supports Unity's WebGL platform. WebGL applications have the advantage of being portable to online websites that can run without any software being downloaded or installed. The portability and accessibility of VO enable VR simulations to be shared online, allowing the observer using the WebGL application to control their perspective of the observation.

Observing can be also done from within the Unity game engine (Fig. 3), enabling data analysis of the reconstruction. We envision this functionality particularly useful for developers, allowing them to identify bugs and usability issues during testing phases of VR applications. Moreover, developer can test their VR experiences on a user-recorded simulation so that, rather than having to frequently re-equip VR headset every time they need to test bug fixes, developers can monitor the reconstructed user actions live whilst focusing on the development output of the Unity game-engine log.

In all of these platforms, the core functionality of VO remains the same in each implementation, enabling playback to rewind, pause and play the reconstructed simulation that is controlled by the observer of the simulation.

4.8 Technical configuration

The live observation was incorporated into Unity's standard Networking configuration for Unity 2018. Unity's default networking handles the transform data of the VR tracked devices for real time (Fig. 5). To register input actions, the VO system monitors for changes in input state of the controllers. When an input change is detected, such as a button state from 'DOWN' to 'HELD', the VR input management class is converted to a JSON string and sent from the client to the server, which distributes this to all other clients connected in the simulation. On the server host, all data received from the connected clients is recorded and saved locally to an XML file once the simulation has elapsed.

5 Methodology

Building upon previous work (Howie and Gilardi 2019), we measure consistency between the data captured on the client computer, server and reconstructed simulation using data captured by the server, showing VO reliability in recording VR user sessions. We are currently only interested in demonstrating the potential for the software for use by researchers, educators and developers to aid VR simulation assessment and development rather than conducting a usability tests (Greenberg and Buxton 2008).

The assessment of consistent data only used the 'Interaction Capture' mode of VO because the environment did not feature any undefined configurations or unknown variable data. A controlled laboratory condition environment was used to validate consistency between client, server, and reconstructed datasets. The observer (one of the authors) hosted a multiplayer session of the simulation (server) located in a separate room next to the laboratory. A wireless local area network (LAN) was used to connect client to the server (observer) machine located in the adjoining room. Due to restrictions of the University firewall, connection between the client and server was achieved using a non-internet-connected router. It is unknown whether this had any negative or positive impact on networking performance or loss of data packets.

Participants for the test were equipped with a VR HMD, two VIVE controllers and three VIVE trackers (for a total of six tracking points) and were asked to start a standalone build of the application on the client machine. After participants entered the local IP address of the LAN server hosted by the observer, they joined the simulation session hosted by the observer. The measured datasets consisted of the client who run the local version of the simulation (Client Dataset, Fig. 7), the server who hosted the multi-player simulation and observed the client in real time (Server Dataset), and reconstruction data generated from an after-action review simulation (Reconstructed Dataset) from the data captured by the server.

Once ethics approval was received by the university, students and staff from the university were recruited via email and word-of-mouth. Five participants volunteered and gave consent to take part. After consent, they were equipped with the VR headsets and allowed to start the session. During the experiment, they were asked to perform the same task repeatedly 35 times, picking up and dropping a sword in a virtual environment. This type of task was chosen as the system captures data only when an input state change is detected. Captured data is guaranteed at points of input state change in client, and, if the system is reliable, these data points should be replicated identically in the server and in the reconstruction, ensuring that the observed interactions and actions in the simulation are as close as they can be to the user interactions and actions. Each participant conducted the test in the laboratory room alone with the observer as the only other (invisible) character within the VR simulation. The observer could move in VR or using the desktop camera to observe the participant from any perspective.

To validate the consistency of the position and rotation data of each tracked VR object, local data captures were recorded on client and server and the server dataset was used to reconstruct the simulation and recapture the data. Local data captures were saved as XML files and were readable by the VO software to reconstruct the actions during a given time action key-frame.

As the client dataset stored inputs and devices states recorded directly from the hardware, it was used as the baseline to determine the consistency of the server dataset. The value d_i, obtained as the absolute difference of values of corresponding data linked to tracker i stored in the server s_i and in the client c_i datasets, as given in Eq. 1, was interpreted as a measure of data consistency between the two datasets; any difference between these datasets was attributed to the live broadcasting of the simulation via the LAN network.

$$d_i = |s_i - c_i|. \tag{1}$$

The difference v_i, computed between the server s_i and the reconstructed r_i datasets, as given in Eq. 2, is to be attributed to our reconstruction system.

$$v_i = |s_i - r_i|. \tag{2}$$

The datasets for the five participants containing the changes in inputs and states for each tracked device were used for the validation of the VO software, the client dataset had in total 4,869 entry points, whilst the server and reconstructed datasets had 4,866. Three packets were lost during the broadcasting between the client and the server; those data were removed from the client dataset during analysis.

6 Results

The difference d_i, as given in Eq. 1, between the client and server datasets and the difference v_i, as given in Eq. 2, between server and reconstructed datasets were analysed by identifying the maximum (worst) difference between the datasets for position and rotation. In the worst case, the client–server difference d_i was of the order of 10^{-7} for position and of the order of 10^{-3} for rotations (recorded as Euler angles), showing that the broadcasting introduced some errors in the data. Despite these small differences between client and server data, we consider the two datasets to be consistent as these small discrepancies are unlikely to be noticed by a human observer. The differences v_i between data in the server and reconstructed datasets were consistently zero for all data points, showing that the reconstruction system preserves the data used for the reconstruction.

A minor loss of packets was noticed when large data was streamed continuously over the network, and happened when a participant left their finger on the controller track-pad during the entirety of the simulation, generating small changes in input states for the track-pad. Considering the large amount of data sent, the loss was negligible in terms of data acquisition with only three instances out of 4,869 (0.0006%) packets lost for all participants data transmitted live. This is likely to be caused by excessive bandwidth used. During instances of packet loss, the system continued to operate using previous data received with follow-up data after the 10ms continuing the live reconstruction. Because the instances of packet loss were only noticed during continuous input state modification, threshold values can in future be used to prevent small changes from registering a state of input change. On a stable and reliable network connection, we do not anticipate that any issues will hinder the live observation for normal input usage once the threshold change value has been included. Because loss of streamed data could be a critical failure in the simulation process, networking features could be implemented to allow for lost packets to be resubmitted to keep the simulations in sync. Alternatively, local client data can be submitted for after-action review if networking issues prevent real-time observation. These limitations are caused by the network transmission of data and out of the scope of this paper.

The reconstruction of tracked VR object motions and input states allow for serial observation of user(s) actions. Rather than restricting analysis to after-action review, live observation allows for instantaneous visual clarification to user's actions and their resulting performance in a VR simulation.

7 Discussion

The validation of the after-action system (Howie and Gilardi 2019) highlighted the potential uses of the VO system, user's actions and motions relative to their real-world posture (Fig. 8) and discovered cases of participants failing to interact with objects, using the wrong input command/button, as well as 'magical interaction' (Bowman et al. 2012), which removes the senses of weight and cumbersomeness from real-life equipment. Participants were also observed to frequently attempt to interact with objects in incorrect positions. During our live observation tests, we noticed similar findings regarding user inputs, but were able to address problems immediately by communicating to the participant using voice communication through an open doorway that connected both rooms that they were pressing the wrong button.

From a developer perspective, the ability to view and replay the input, actions and movement of users offers the chance to directly observe the process that lead to bugs or errors, as well as observe human behaviour during simulations. The problem with noticing issues during and after-action review is that the user has already completed the simulation by the time identifiable issues are found. If the user is observed live in real time, as allowed by the system in the paper, issues can easily be noticed and resolved without the user struggling and becoming frustrated by the lack of guidance caused by interaction or usability issues, which can be especially true for novice users (Tromp et al. 2003).

For simulations to be replicated as experienced by the user, the same build version of the application must be used during the live observation or after-action review reconstructions. If a different build version is used to reconstruct user data, it may result in the motions and actions of the reconstruction not replicating the same experience as the user. The need for consistent build versions only applies to situations where the changes from the reconstructed simulation build have a direct impact on the user interactive experience. For example, if a user picked up a box in a dataset (Build v1), but the box was no longer present in the updated version (Build v2), the actions and motions of the user would remain during reconstruction but no context would be available to understand what is happening from the observers perspective as the box is no longer exists. Therefore, backups of the published simulations should be archived to ensure when reviewing participant performance, the reconstructed simulations are not impacted by modified conditions.

In this study, we used the 'Interaction Capture' mode of VO for the recording and reconstruction of data. It is important to highlight, however, that any type of tracking

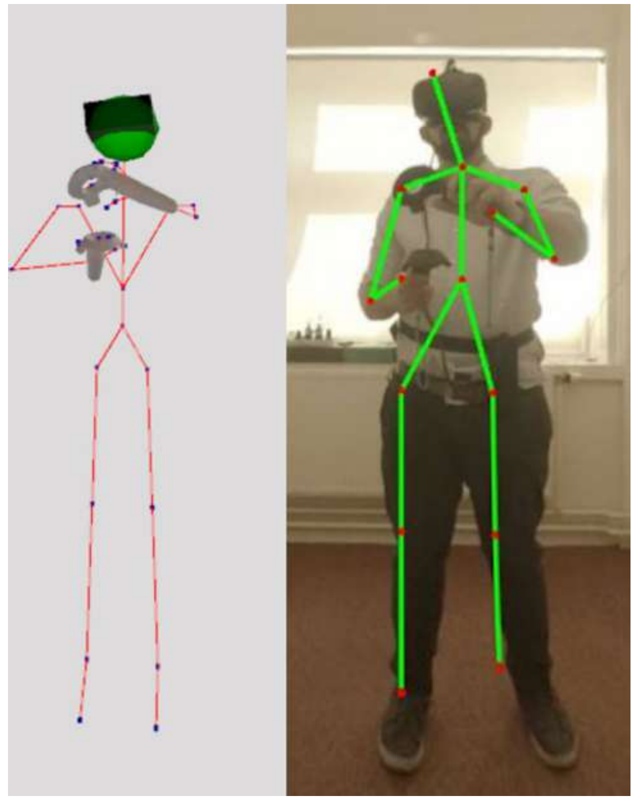


Fig. 8 Side-by-side comparison of an estimated skeleton posture obtained from the VO system (left) and a participants posture from GoPro video footage with OpenCV's pose estimation (right). The skeleton posture will deteriorate as and when the participant moves untracked joints of their body. In Howie and Gilardi's case, the lack of torso tracking caused the entire structure to rotate to match that of the HMD. Because feet and knee joints were untracked the IK system estimated the position of the knee joints relative to the ground and height of the HMD. Therefore, if the participant moved their feet from the default standing pose as seen above, the lower parts of the body would stay rigid

Fig. 9 VO was used in a commercial project for remote observation of a VR simulation during development giving the client the opportunity to monitor progress using VO WebGL and Desktop builds of the VR application. VO was later used to review trainee VR experiences, as seen here: https://www.dropbox.com/s/uaakovec3518cvx/VirtualObservationVRDemonstration.mp4?dl=0

device that supports the Unity object system for updating the transforms of the tracking device(s) will be supported by the system. This functionality aims to offer universal support to suit the needs of researchers and developers working with novel technology. To demonstrate the flexibility of the system, we used a Leap Motion hand tracking device to capture the object references our hands in Unity, whilst we interacted with a virtual car desktop screen to modify the car configuration. The movement of the hands whilst interacting with the virtual desktop screen was reconstructed, repeating the motions and button presses conducted during the recording (Fig. 3). This type of functionality can extend to other areas of tracking devices such as Face Tracking (Li et al. 2015; Olszewski et al. 2016) and Eye-Tracking (Meißner et al. 2017) which can offer greater insight into user experience. Jacob and Karn (2003) have stated that with eye-tracking, the scan path, gaze interest and fixation interest can indicate a user's intended target before they could 'actuate any other input device' (Jacob and Karn 2003, p. 589]. The inclusion of VO software will allow for these interaction devices to be saved and reconstructed, improving the data collection and analysis for studies. Using the VO software, it can streamline the process of assessing users, with reconstructed simulations able to accommodate several tracked devices and reconstruct them simultaneously with relative contextualisation to the other devices in the simulation.

In Howie and Gilardi (2019), the skeleton posture created using Unity's IK system was assessed, concluding that head and hands were tracked accurately with data provided from the VR equipment, but torso and feet tracking proved lacking since no real-world object was used to calibrate their position or orientation (Howie and Gilardi 2019). In this study, we rectified this issue by using three additional trackers, attached to both feet and the torso. Using all six tracking positions (HMD, two controllers, two feet trackers and torso tracker), we were able to achieve 'full-body' tracking of users in real-time observation and after-action review simulations. This tracking set-up could be used to monitor a user's location of core body points relative to the local tracked physical 'play-space' they have configured for their VR set-up.

Developers and researchers can use VO to analyse feasibility and usability of systems. Educators that adopt VR in their classrooms can use the data to check and assess how users have completed training procedures and verify they are correctly handling equipment or safety procedures. In Howie and Gilardi (2019), it was highlighted that participant data recorded of a fifteen-minute simulation using the VO system was always significantly smaller (around 80MB) than equal length video recordings using GoPro equipment (around 5.8GB). This continues to be true, and we demonstrated that the data can be transmitted on a network to obtain a real-time live simulation of what users are doing within a VR simulation.

To test the portability the VO system, we exported the project to a separate Unity development project which we used to create a WebGL application that replayed the data recorded during one of the validation sessions (see Fig. 6 for controllable observation demonstration). Because SteamVR does not support WebGL, VO bypasses these restrictions enabling the VR simulation to be reconstructed and observed in web browsers, which animation capture of simulation data for input modification would be unable to achieve due to the lack of SteamVR compatibility (see example in Fig. 6). The examples presented in this paper show the flexibility of VO for recording VR user simulations and reconstructing them on multiple platforms.

Datasets recorded from users with the VO system can be used to replicate the actions and motions of the user, simulating the use of VR equipment without relying on the equipment resources. The ability to share the user sessions allows external observers to analyse datasets without having access to the equipment, software and development area of the project and allowing researchers (after ethical approval and user consent) to share data with collaborators and reviewing panels, as well as allow other researchers validate research findings. Using our system, software houses can report bugs accurately to the relevant department and educators to document trainee assessments.

An issue we faced during this study was the variation of positional data of tracked objects during the action of picking up and dropping objects. We originally intended to compare the transform of each tracked device during the instance of the pickup or drop action being completed; however, we noticed slight inconsistencies in positional and rotational data between the datasets. After investigating the data recorded on the client PC and server PC, we discovered that the issue was caused by the minor modification of the controllers in real space during the fractional time difference of input state recorded and an action being captured. We anticipate this minor discrepancy in data was caused by the real-time updating of the render models which aims to keep the virtual controls consistent with their position in real world to satisfy the user experience. This minor discrepancy in pickup position does not affect the VO system as the action transform positions were consistent for both the client and server datasets. To keep action and input persistent during reconstruction the transform of the interaction devices were modified prior to the physics update of the game engine.

8 Limitations

Packet loss during the transmission of data could result in loss or corruption of simulation data as an action-frame would be skipped from the reconstructed simulation during live observation. Like any set of data submitted over a wireless connection, there is potential for individual packets data to be lost. Unfortunately, this is a networking issue with the Unity game engine networking architecture and not our VO. We have provided potential solutions in Sect. 6, to mitigate this issue.

Recording of non-serialisable Unity components requires the developer to extend within the source-code the required variables, potentially making the 'Full Simulation Capture' mode too complex for non-technical users. We hope to improve this in the future, making the 'Full Simulation Capture' mode as user-friendly as the standard mode. One of the difficulties faced with the Full Simulation Capture version of the software was the inability to record data of the additional tracked objects dynamically. We were able to successfully record all data (variables) from objects and reconstruct them using system reflection. However, system reflection has an unavoidable impact on system performance and caused too much lag in the simulation during reconstruction.

9 Conclusion

This paper built on Howie and Gilardi's Virtual Observation system (Howie and Gilardi 2019) for observing the motions, inputs and actions of users in virtual reality simulations. VO was validated in Howie and Gilardi (2019) for after-action review of users performance in a detailed fire training simulation scenario (Fig. 1), proving that the VO system was capable of reconstructing detailed simulation from only the input and actions of a participant with no other tracking data required. In this paper, we have demonstrated the capability of the VO system for reconstructing simulations in real time with consistent datasets for live and after-action (offline) review of user sessions, capable of both full body tracking and hand tracking set-ups. We also demonstrated the ability for researchers, developers and educators to share VR simulation data to conventionally unsupported platforms (WebGL).

We conclude that these recordings are on par with Lopez et al. (2017), but offer greater clarity into high-level technical details of the simulation and the role played by the VR user (how the actions of the user affect the simulation procedure), offering context from cause and effect relationships of user's actions (Hanoun and Nahavandi 2018). With the inclusion of live observation, user performance can be immediately assessed with the potential for observers to have instant clarification to issues the user may be experiencing. For research purposes, the ability to generate visual data output to contextualise VR research can also be a key benefit for researchers and study validations. Given that screenshots and videos can often have limited contextual awareness, the ability for researchers to control the observation of user data themselves can help clarify authors' findings.

VO was incorporated into a commercial product to let clients review the functionality of a VR simulation through a WebGL visualiser without need for software to be downloaded or without the constraints of pre-recorded videos. Using the VO software, the client was able to view for themselves the interactions and protocols of a training simulation remotely from the development location without requiring VR equipment or additional software. Unfortunately, due to confidentiality agreements, we are not able to share any additional data or details at the moment (Fig. 9).

For researchers, the VO system offers greater scope for experiment complexity and reach, allowing for studies to be conducted in external locations ('in the wild') with data retrievable for analysis. For development of VR applications, this software can be of use to allow identification and replication of bugs with consistent repeatable simulations from a single capture dataset to ensure reliable analysis is conducted. After resolving a bug or issue experienced by the user, the same user dataset can be rerun to ensure that the issue captured originally is not repeated, indicating when the issue is resolved.

We conclude that VO of reconstructed simulations is a highly effective and versatile tool for researchers, educators and developers which is enhanced by the ability to observe simulations in real time. The streaming of data in real time during live observations also circumvents the need for data to be stored on external services, avoiding the issues faced by bandwidth usage experienced in Howie and Gilardi (2019).

For future work, it will be interesting to determine whether and how the VO system could be incorporated directly into an automated evaluation system that provides instant feedback (Hanoun and Nahavandi 2018) for situations where an observer is not available.

Acknowledgements The authors would like to thank Swagelok Scotland Ltd. for using our VO system to review simulation procedures, protocols and interactions during development of their training application and allowing the screenshot from their application to be included in this paper.

References

Blomberg J, Burrell M, Guest G (2007) The human-computer interaction handbook. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, chap An Ethnographic Approach to Design, pp 964–986. http://dl.acm.org/citation.cfm?id=772072.772133

Bowman DA, McMahan RP, Ragan ED (2012) Questioning naturalism in 3d user interfaces. Commun ACM 55(9):78–88

Carranza J, Theobalt C, Magnor MA, Seidel HP (2003) Free-viewpoint video of human actors. ACM TOG 22:569–577

FitzGerald E (2012) Analysing video and audio data: existing approaches and new innovations. Surface Learning Workshop 2012, Bristol, UK

Goldberg SL, Knerr BW, Grosse J (2003) Training dismounted combatants in virtual environments. Technical report, Army Research Development and Engineering Command Orlando FL Simulation

Greenberg S, Buxton B (2008) Usability evaluation considered harmful (some of the time). In: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, pp 111–120

Greenhalgh C, Purbrick J, Snowdon D (2000) Inside massive-3: flexible support for data consistency and world structuring. In: Collaborative virtual environments: proceedings of the third international conference on collaborative virtual environments, vol 2000. CiteSeer, pp 119–127

Greenhalgh C, Flintham M, Purbrick J, Benford S (2002) Applications of temporal links: recording and replaying virtual environments. In: Proceedings IEEE virtual reality 2000. IEEE, pp 101–108

Hanoun S, Nahavandi S (2018) Current and future methodologies of after action review in simulation-based training. In: Proceeding of the 2018 annual IEEE international systems conference (SysCon). IEEE, pp 1–6

Howie SR, Gilardi M (2019) Virtual observation of virtual reality simulations. In: Extended abstracts of the 2019 CHI conference on human factors in computing systems, association for computing machinery, New York, NY, USA, CHI EA'19, pp 1–6. https://doi.org/10.1145/3290607.3312836

Jacob RJ, Karn KS (2003) Eye tracking in human–computer interaction and usability research: ready to deliver the promises. In: The mind's eye, Elsevier, pp 573–605

Jung B, Amor HB, Heumer G, Weber M (2006) From motion capture to action capture: a review of imitation learning techniques and their application to VR-based character animation. In: Proceedings of the ACM symposium on Virtual reality software and technology. ACM, pp 145–154

Lang P (2019) Final IK. https://assetstore.unity.com/packages/tools/animation/final-ik-14290. Accessed 04 Apr 2019

Lazar J, Feng JH, Hochheiser H (2017) Research methods in human–computer interaction. Morgan Kaufmann, Burlington

Li H, Trutoiu L, Olszewski K, Wei L, Trutna T, Hsieh PL, Nicholls A, Ma C (2015) Facial performance sensing head-mounted display. ACM ToG 34(4):47

Lopez T, Dumas O, Danieau F, Leroy B, Mollet N, Vial JF (2017) A playback tool for reviewing VR experiences. In: Proceedings of the 23rd ACM symposium on virtual reality software and technology. ACM, p 83

Meißner M, Pfeiffer J, Pfeiffer T, Oppewal H (2017) Combining virtual reality and mobile eye tracking to provide a naturalistic experimental environment for shopper research. J Bus Res 100:445–458

Olszewski K, Lim JJ, Saito S, Li H (2016) High-fidelity facial and speech animation for VR HMDs. ACM TOG 35(6):221

Petrie H, Hamilton F, King N, Pavan P (2006) Remote usability evaluations with disabled people. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, pp 1133–1141

Reitmayr G, Schmalstieg D (2005) Opentracker: a flexible software design for three-dimensional interaction. Virtual Real 9(1):79–92

Tromp JG, Steed A, Wilson JR (2003) Systematic usability evaluation and design issues for collaborative virtual environments. Presence Teleoper Virtual Environ 12(3):241–267

Von Spiczak J, Samset E, DiMaio S, Reitmayr G, Schmalstieg D, Burghart C, Kikinis R (2007) Multimodal event streams for virtual reality. In: Multimedia computing and networking 2007, international society for optics and photonics, vol 6504, p 65040M

Wobbrock JO, Kientz JA (2016) Research contributions in human–computer interaction. Interactions 23(3):38–44

Appendix G

Papers Submitted for Review

Noname manuscript No.
(will be inserted by the editor)

Design and Development of Authentic Assessment Tools for Room-scale Virtual Reality Training

Scott Howie · Marco Gilardi

Received: date / Accepted: date

Abstract Virtual Reality has the potential to change the way industrial training is performed, allowing trainees to experience scenarios that might be dangerous or expensive to recreate in real life. Although a large body of literature presents such experiential learning approaches, little is understood of how authentic assessment can be implemented in VR training. In this paper, tools for authentic assessment and contextual feedback are designed, integrated into a streamlined framework, and evaluated within an industry-based scenario. The tools were developed based on findings from the literature and from a contextual user study with industrial assessors. To enable the framework and the tools to work, a novel dual-recording methodology that allows to capture all data within the simulation without affecting the frame-rate was developed and implemented within the framework. The qualitative evaluation conducted with academic and industry assessors compared the framework against video recording and analytical data forms of assessment for the same room-scale VR training simulation, showing that the framework is capable to support authentic assessment and contextual feedback better than the other two approaches to assessment.

Keywords Virtual Reality · Tools · Authentic Assessment · Feedback · Training · Dual-Recording

1 Introduction

With the advancement of virtual technology, assessment capabilities for virtual training are undergoing changes, with advances in technology driving the ability to capture authentic, complex and skill-based assessments [47]. However, unlike real-world training, virtual training alters the context in which trainees are observed, evaluated and assessed on their performance against the learning outcomes of the training simulation. As discussed by Howie and Gilardi [23], performing contextual observation of trainees in VR is challenging, as the trainee and the assessor are positioned in two different contexts, the assessor is in the real-world whilst the trainee is in the virtual context. Observing the trainee from inside the same VR simulation allows insight into their performance [21] and helps in determining the cause and effect relationships from user actions [21]. Observation of trainees from their point of view can offer insights and circumvent issues often encountered of trainees describing inaccurately what they did due to a lack of awareness or understanding of the task or system under study [5]. However, direct perspective observation can induce motion sickness on the observer [37], and the presence of additional recording equipment could cause unnecessary pressure and nervousness [45].

Virtual forms of training are useful for learning since they provide an objective platform for assessment that can be tracked and built-upon to improve learning without risks to the trainee [33, 8, 53]. At present, training simulations often rely on assessment based on Artificial Intelligence (AI) systems to replace assessors [34], missing out on a wealth of real-world experience assessors can provide to trainees through feedback. Automatic assessment of complex skills or rich data is not yet viable and may still take years be on par with assessment and feedback from assessors [47, 3, 1].

Assessment within VR training simulations can be designed to ensure that tasks and objectives defined are true learning outcomes which measure the trainee performance and competency [53]. As such, it is important that feedback follows assessment, referencing the trainee performance [53]. Although assessment and feedback in simulations are sometimes considered a time-consuming by assessors and trainees [53, 51, 16], these tasks are invaluable for ensuring that learning outcomes from the training simulations are achieved. To mitigate time-consumption in assessing a training performance computer system that aid assessors during the after-action review process should be used [29, 21].

Assessment is important in VR training simulations as it provides evidence of the trainees current knowledge [55, 56, 30, 22], which can be used as a platform for providing feedback. Pedram et al. [50] findings on trainee learning outcomes using an immersive VR training simulation suggest that assessors had a direct impact on the learning outcomes of trainees, concluding that explanatory feedback has a positive outcome on learning process, stating that it should be considered an essential component for any training program [50]. These findings are further validated in the literature, with Martins and Kellermanns [41] and Wan and Fang [62] presenting similar arguments. Pedram et al. [50] also note that that trainees can learn and benefit from self-reflection of past training experiences and useful feedback provided by skilled assessors. Johnson and Priest [30] argue that for novices, explanatory feedback is more effective than corrective feedback, as explanatory feedback provides the reasons behind their success or failure, whereas corrective feedback only informs of a success or failure [50].

Two-ways communication between trainee's and experts allows for opportunities to correct mistakes and misunderstandings, and further improve learning in trainee knowledge [50]. Furthermore, Pedram et al. [50]'s findings are in line with those of traditional training [33, 8], suggesting that traditional and virtual methods of assessment and feedback are closely related.

The ability of VR training simulations to reproduce real-world situations lend itself to authentic assessment approaches. Newmann and Archbald [44] describes authentic assessment as a tool to provide high-order thinking and problem solving capabilities for the individual and society. Authentic assessment approaches trainees assessment based on their performance on the completion of intellectual tasks that resemble tasks they will encounter in real-world situations, rather than on abstract and indirect performance indicators [63]. According to Wiggins [63], simulating the real-world to provide authentic assessment that reproduces life-like capabilities is the main essence of authenticity. Conducting authentic assessment involves the use of authentic tasks that are relevant to the context that determine student learning outcomes [20], this is often because jobs focus on authenticity of performance and competencies [19]. Unlike assessment of authentic achievement, where emphasis is placed on the learning outcomes, authentic assessment examines the authenticity of how the assessment is conducted [32]. However, performance demonstrations between contexts are not indicative of each other, as the domain knowledge could be contextual to a relevant learning scenario [32].

Authentic assessment can be achieved in VR training simulations by measuring the application of real-world knowledge during meaningful training and assessment exercises [64, 36]. The ability to replay and review trainee performances is an important tool for such assessment [35], and provides a assessor the ability to observe and monitor trainee performances.

This paper contributes to the field of virtual reality training simulations by presenting tools for authentic assessment organised in a framework that enables assessors to conduct authentic assessment and generate meaningful feedback of VR trainee performances. This framework was designed to meet the needs of oil-and-gas industry training experts, allowing unrestricted observation and assessment of trainee VR performances with the ability to generate contextual and targeted feedback to trainees anywhere in the world. Although the initial design was targeted to industry experts the framework is sufficiently general to be applied in other domains, as suggested by the evaluation conducted with industry and academic experts. The evaluation of the framework implementation against existing methods of assessment for analytical data and video, demonstrated the advantages offered by our approach. Furthermore, the assessment and feedback tools streamlined authentic assessment operations, lessening the cognitive overload for the assessor.

2 Related Work

Many forms of assessment for VR simulations exist, such as analytical data [27], environmental recording [15], video recording [12], motion and input capture [38], stealth assessment [57], action based [11], and animation [37]. In this work we focused on evaluating and comparing the first three methods mentioned in the list above, namely analytical data, environmental recording and video recording. In this section related work to these three methods of assessment in VR is presented.

2.1 Analytical Data

Analytical data gives simulations the ability to provide a consistent platform for assessment [27]. Analytical data captured in a simulation can vary in scope and complexity, and can be used to evaluate the simulation outcomes. Examples of such data can be time of completion, number of errors, motion data, objectives completed, as well as additional variable data that are unique to the training simulation [34, 33, 27]. Analytical data may be captured automatically by the software experience [27, 7], or require a assessor to record the data and score parameters [14]. The measurement of time as a metric is widely used when judging trainee performance, as it is easy to compare against a target time value [8]. Analytical data can be used by automated assessment systems to provide instant and automatic outcomes for the trainee performance [27]. In the most primitive form of analytical feedback, the data collected from the simulation can be used to present the performance analytically using tables and graphs or simple pass or fail outcomes [8], which are insufficient for effective evaluation [48]. In these instances analytical data methods are useful for creating generalised outcomes of trainee performances, notably in the ability to identify novice and expert assessors [27] based on the score differences.

2.2 Environment Recording

Environment recordings are a form of recording which capture variable information within a game or simulation. With environment recordings, the focus is capturing sufficient data that would enable relevant sections of the environment to be captured for reproduction [15, 61, 52]. Knowing that external assessors are capable of conducting manual assessment of trainee performances [21], environmental recordings appear an effective method of visually capturing and reproducing the trainee performance. Environmental recording improves upon animation based recordings by offering more flexibility in the approach for recording data [37], demonstrating an effective method of observing training simulations.

One of the earliest examples of capturing simulation environments and experiences is Goldberg et al. [15]'s work, which used early implementations of VR technology to replay and review army training simulations in an effort to measure users performance. Goldberg, Knerr and Grosse's [15] after-action review system offers a guarantee that events can be triggered and reviewed repeatedly, but the focus of the recording is on the event data of the training performance, not on the trainee actions. This focus makes difficult to identify any usability issues experienced by the trainee since their input is not replicated. Understanding trainee usability issues is a key requirement of understanding the trainee performance, especially during remote VR training sessions.

A similar approach to [15] is taken in Raij and Lok [52]'s IPSViz an after-action review tool for human-virtual human experiences that aims to reproduce the simulation experience with a virtual-human, overlaid with tracking data, audio, video, speech, event logs and human behavior. Raij and Lok [52]'s approach is a hybrid model that overlays on-top of multi-view video and audio data, 3D reconstructed rendering, sensor data and transcripts of conversations to provide in-depth contextual knowledge of the real-world posture and location of the user relative to the virtual simulation data. The IPSViz system is designed to enable self-evaluation and review, allowing the trainees to gain insight into their performance to improve future work [52]. However, IPSViz focuses on data presentation rather than reconstruction of sophisticated simulation experiences.

Another after-action review approach to environmental recording is Stone et al. [58] 3D underwater mine countermeasure training simulation. To capture trainee data for assessment, Stone et al. [58] simulation logs trainee data that is time-stamped with their progress exploring the underwater virtual environment, including other search parameters for dwell times and identification of objects. To conduct after-action review, a summary screen presents high-level performance indicators of the trainee, highlighting the summative information for objects identified and hit percentage [58]. After reviewing the summative data, the performance could then be replayed using a non-linear reconstruction methodology which enabled a visual observation of the trainee performance by the assessor [58]. The replay tool included embedded feedback tools which enabled the assessor to highlight on the screen by virtually drawing onto the reconstruction while discussing the feedback with the trainee. Furthermore, the tool includes the ability for the assessor to view the trainee from multiple perspectives, including a free-camera control which offered the assessor full control of the observation and provided an assisted feature that reproduced the search path of the trainee during the die [58].

Howie and Gilardi [23] develop a virtual observation system able to capture the trainee VR simulation for after-action review, providing tools for play, pause rewind and skip to time. However, Howie and Gilardi

[23]'s virtual observation system is limited as the capture of data was restricted to events during action and motion key-frames for serialisation, required manual assignment of objects for recording, and when attempting to record too much data, the system performance was impacted, reducing the frame-rate of the simulation.

In comparison to current room-scale virtual reality training simulations which would require significant capture of variable information necessary to enable authentic assessment of the trainee performance, IPSViz emphasised communication methods such as voice, audio, gaze and kinetic movement [52] and Stone et al. [58] focused on search paths and detection of objects [58]. As such, both IPSViz [52] and Stone et al. [58]'s assessment methods enables effective evaluation of trainee performances for their intended purposes, but due to their incompatibility with modern simulations they are not designed to capture and reconstruct room-scale VR training simulations [15, 17, 61].

2.3 Video Recording

Video recording is capable of recording data without being directly embedded within the simulation environment. Video and audio recordings are a common way for laboratory based and 'in the wild' studies to document experiments and procedures that can be qualitatively analysed at a later stage[12]. Video recordings make it possible to analyse the the performance of a trainee simulation through direct visual observation, and are stored in a format that is easy to share between collaborators [12].

Video recordings assessment is a useful and easy to implement method that allows a recording of a training performance to be captured for assessment by assessors, with non-linear functionalities [54]. Video recordings can quickly be skimmed and timestamps can be used to identify and mark points of interest from the trainee performance [26]. During assessment the video recordings can be paused, replayed, edited and zoomed in or out to focus on specific observation areas [13]. The use of video recordings of VR simulations [49, 21] is due to the prevalence and accessibility of video recording software and its compatibility [12]. Furthermore, video recordings can be used to capture the simulation, and the real-world movement and interaction of the trainee operating the VR equipment [35].

However, video recordings may be insufficient to identify constraints in the trainee performance related to VR room-scale movement. For example, objectives outside the reach of the trainee physical movement space may result in a trainee failure, which is due to lack of ability to reach rather than the trainee lack of knowledge [24]. A potential solution to this problem is to capture both virtual and real-world using video recordings of the trainee performance for assessment, generating context for both environments [24]. However, the presence of a video camera in real-world could alter training outcomes, as the recording equipment may cause unnecessary pressure and nervousness on the trainee [45]. Face-to-face training would also put a strain on the trainee due to the presence of an assessor in the same space as the trainee, but it could be argued that it would be less intimidating knowing mistakes or actions could not be replayed or seen by anyone else. Video recordings as a method for capturing data limits the quality and quantity of data that can be acquired by the simulation for the purposes of assessment [18].

3 User Study

A qualitative contextual study to identify the requirements for the framework for authentic assessment of room-scale VR simulations was conducted with five oil-and-gas training experts that work in a Scottish company, this number of participants was judged as sufficient according to Malterud et al. [40] information power theory. The contextual study used semi-structured interviews to: (a) identify the current practices in classroom-based industrial training assessment and feedback, (b) discover features or functionalities for a virtual assessment and feedback framework, and, (c) identify any concerns about moving to a virtual medium for conducting assessment and providing feedback. Allowing the researcher to explore interesting elements of the conversation with the assessor during the interview.
Transcripts of the semi-structured interviews were analysed by the first-author of this paper using thematic analysis following Braun and Clarke [6] process. The identified themes were then independently reviewed by the second author and discussed before finalising them. From the thematic analysis of the five semi-structured interviews and on-site observations, the following final themes were identified: **observations** (see Figure 1), **framework features** (see Figure 2), and **concerns** (see Figure 3).

3.1 Observations Theme

The observation theme identified several sub-themes which created overarching topics of discussion regard-

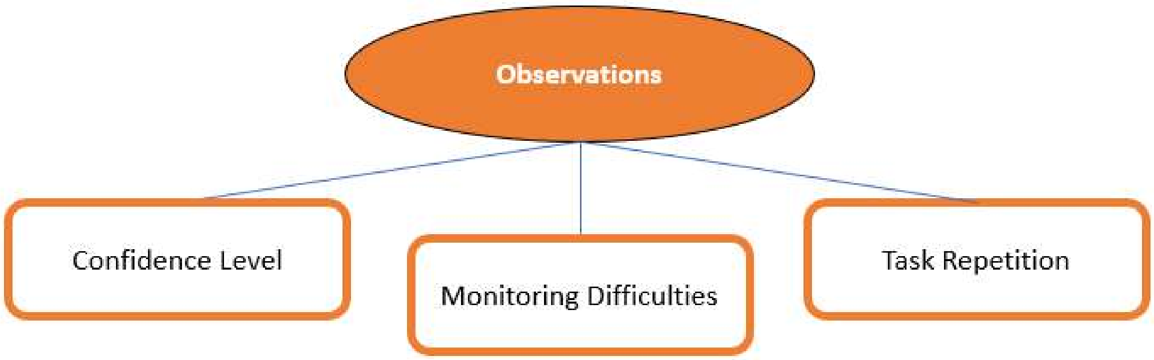


Fig. 1: Themes and sub-themes for observation.

ing trainee observations, assessments and procedures for conducting assessment, see figure 1. The observation theme focused on discovering how trainee performances could be reproduced in a format that is usable and useful for the assessors to directly observe and monitor trainee performances. One of the interesting findings from the observation theme was identifying how participants judged trainee's during their assessments in classroom-based training, highlighting that confidence level played a key-role in determining how the trainee was anticipated to perform. Trainees who the assessor felt were likely to struggle were given additional attention during the assessment to quickly identify anticipated failures.

The following sub-themes describe the process of how observations are handled within a oil-and-gas training company, providing information to the approaches and methodologies assessors adopt when monitoring trainees and how they judge trainee knowledge. By observing the trainee during assessment, participants re-enforced the idea that the focus of assessment is to determine the knowledge of the trainee, and, by directly witnessing them logically apply knowledge to a set of tasks, the assessor is able to determine how satisfactory they performed.

Trainee confidence level was one of the sub-themes identified and plays an important role in how assessors judge trainees during assessments in classroom-based training, highlighting that the trainee confidence level played a key-role in determining how the trainee was anticipated to perform by the assessor. Trainees who the assessor felt were likely to struggle, were given additional attention during assessment and training sessions to monitor for expected mistakes. Trainee confidence was measured by identifying the movement, actions and self-directed comments of the trainee to determine if they are applying knowledge logically. Participant 3 observes:

> *"If somebody dives in that's an immediate this guy's just a bit cowboy, doesn't really know what he's doing as opposed to someone who's thinking things through first and you can see them having a look around, reading instructions properly, before they've even gone anywhere near the hardware. That's a big difference and usually if someone is doing that and you can see them thinking it through it's a good indication." (Participant 3)*

While confidence levels did not provide a definitive indication regarding a trainees pass or failure in the assessment, it provided the assessors with an opportunity to identify a participant that is lacking in knowledge in a specific area, or to speak to a trainee who is approaching the tasks too fast and not considering the health sand safety risks attached.

Monitoring difficulties was the second sub-theme identified. Albeit infrequent, monitoring difficulties can occur in classroom-based assessment. Participants suggested they may have missed an action, movement or identifier in the trainee performance. Participants reflected that the challenge of keeping track of the trainee progress, was due to the loss of focus when diverting attention between other trainees in the class, potentially causing issues to be overlooked, as commented by Par-

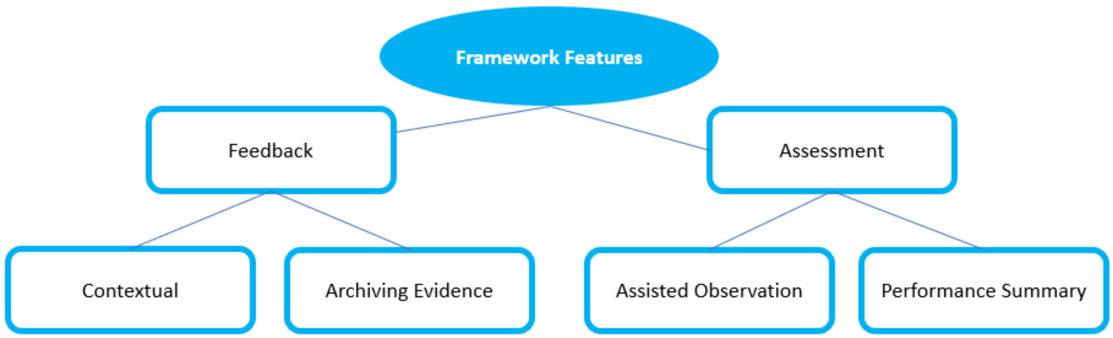


Fig. 2: Themes, sub-themes and sub-sub-themes for framework features.

ticipant 5:

> *"You can't concentrate on four people all the time. If they're all doing different things you've only got one set of eyes and you can't really keep your eye on everything at the one time." (Participant 5)*

Therefore, while a assessor is monitoring one trainee completing an objective, another trainee may have made an error that goes unnoticed. In this case, the error would not be identified until the very end of the assessment when the practical work is dismantled and examined. Even so, participants noted that there is no way of identify if the trainee miss-used equipment, preventing an opportunity for feedback and to correct their gap in knowledge.
Findings also suggest that it is best to conduct assessment as a one-on-one operation to maximise the learning opportunity potential for the trainee to target the weakest aspects of their performance. However, this one-on-one approach is not feasible in commercial classroom-based training due to economic factors that involve the cost of facilities, staff and the length of each assessment. One assessment observed during field-observations lasted the maximum available time of two hours, requiring an assessor to be present for a single trainee conducting the assessment.

Finally, *task repetition* was the final sub-theme identified, and it enables assessors to efficiently conduct the assessment by analysing trainee performances using their experience and expectations from previous assessments. The task repetition of the trainee performance during training or assessment provides a gauge for how well the trainee is coping with the challenges. The high probability of failure at certain points of an assessment influences assessors observation, causing their attention to be diverted to focus on a trainee when they reach an area of the assessment that is prone to causing failures. However, with each assessment given a set criteria and running time, the assessor has to be present for the entire duration, meaning there is a lot of down-time waiting for trainees to complete tasks. As such, assessors expressed the desire to skip sections of the trainee assessment where the trainee is preparing, by reading the instructions and familiarising themselves with the training environment.

3.2 Framework Features Theme

The framework features theme was identified during the analysis and provide insight about the features and tools required for the assessment and feedback within the framework. This theme comprises of two sub-themes, each divided in their own sub-sub-themes, see figure 2.

3.2.1 Assessment Features

Assessment features was the first sub-theme identified under the 'framework features' main theme this theme highlighted the needs of participants during the assessment of a trainee performance. This theme is composed by the following sub-sub-themes.

Assisted observations was seen as one of the main benefits to virtual forms of assessment, allowing the assessors observe trainee in the simulation to support assessment to reduce the probability of overlooking data and aiding assessors overall ability to identify trainee issues. The ability to review the trainee session was seen as an option to provide better context for feedback, allowing the trainee to review their recorded performance, and allowing easy identification for sections of the trainee performance that the assessor could reference in their feedback.

As discussed in section 3.1 all of the trainees conduct the assessment in the same room at different speeds based on their individual skill-levels, as such, it is easy to lose track of what stage and task each trainee is on. To avoid loosing track *action identification* is fundamental for the assessment. assessors need to make sure that they do not overlook any of the actions performed by the trainee. To achieve this assessors would position themselves in an area of the room to observe all trainees, without intruding on their work space. As assessment is often required for jobs and certifications, most trainees have some anxiety or nervousness. Because of this, assessors do not want to observe the trainee too closely as it could cause unnecessary pressure on the trainee, causing them to panic and make a mistake. assessors reflected that an improvement for activity monitoring of trainee actions would be to categorise each section of the performance to focus on one task or objective, the assessor would then be able to quickly and easily identify the important or abnormal areas of the the trainee performance.

> *"I think if you see somebody who is reading the sheet all you need to say is they've read it. You don't need to watch them reading it. They might take 10 minutes to read the instructions. I don't need to watch them reading it for 10 minutes. All I need to do is say tick they've read it and then if I can skip through to the next action and observe that, so I think if the program was able to identify an action and it would be that there was a trigger in there that says that's the next action then I think that would be safe." (Participant 4)*

Participant 4 highlights that assessors could use an action identification system to determine when a trainee is conducting assessment tasks and when they are reading instructions. With this information the assessor would skip to only the relevant parts of the assessment where the trainee is conducting actions that are part of the marking scheme or for spotting irregularities in a trainee performance. These features provide quick insight into the overall actions conducted by the trainee. However, assessors stressed that an action identification system to identify and categorise different sections of the trainee performance, would not be a replacement for directly observing the trainee, and would only be used to assist the assessment operation. By using a remote virtual form of observation, from a multi-view perspective, the trainee would be unaware of when and where assessor watching. Such features would provide assessors with the ability to focus their attention to specific areas of a trainee performance.

Performance summary Assessors were found to focus on specific elements of a trainee performance relative to their task on during the assessment. When monitoring a trainee, the assessor relied on their expectations and experience observing the assessment processes to determine the objective progress of the trainee for passing or failing each objective. As such, a performance summary would allow assessors to see an overview of the trainee's outcomes, which could be used to identify the immediate issues of trainee performance that require intervention by the assessor, rather than being restricted to passively observing the trainee until a issue is visually observed.

3.2.2 Feedback Features

Assessment features was the second sub-theme identified under the 'framework features' main theme this theme highlighted how participants create and provide feedback to the trainee. This theme is composed by the following sub-sub-themes.

Contextual feedback is seen as crucial to the learning experience when providing feedback to the trainee. assessors stressed the importance of supplementing the feedback with sufficient information to replace the use of traditional in-person gestures, which use a combination of pointing and other visual cues to identify and focus the attention of trainees. When a trainee struggled to understand an objective the assessor frequently points to each individual element of the process, followed by verbal explanation, demonstrating the steps to a trainee. Likewise, during trainee practice work, assessor highlight and correct mistakes by guiding trainees attention to mistakes by pointing and explaining the trainees error. The gesture is intended to focus the gaze of the trainee to a specific area of the work environment. The use of gestures is considered a method of grounding speech to a physical environment, clarifying meanings from verbal messaging to physical constraints in the environment, informing the trainee what to focus their attention on [60]. As such, it is important that

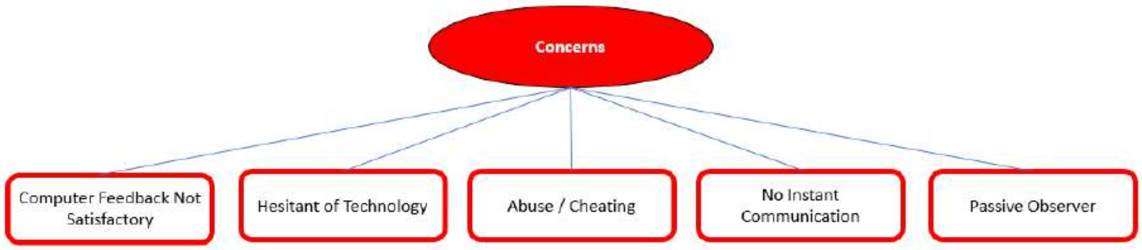


Fig. 3: Themes and sub-themes for concerns.

the feedback supplied is clear and concise, defining exactly what the target of the feedback is and providing a contextual presentation to the trainee. The feedback must be able to stand on its own, with clear information that is understandable and useful for the trainee to digest and learn from. When referencing sections of the trainee performance for feedback, the assessor should make full effort to contextualise the feedback in a manner that is understandable. However, given the complexity of training simulations, it may be too difficult to rely on verbal feedback alone, therefore, it may be necessary to supplement the feedback other elements, such as video of images of the objectives or equipment.

> *"[For feedback] you could overlay [voice]. I suppose you could watch the video and overlay voice over it after the event or you could type annotations or something [into] the video." (Participant 3)*

As discussed by Participant 3, by generating feedback that references the trainee performance within the simulation, assessors would be able to clearly and concisely describe the mistakes made by the trainee using the visual cues, while the trainee is able to re-witness their actions, receiving relevant feedback to support their leaning.
When trainees refuse to accept failure outcomes, escalating the case to an external reviewer and challenging the assessment judgement of the assessor, *archiving evidence* plays an important role in classroom-based scenarios. The capturing and archival of evidence in preparation of disputes is occasionally required. Trainee mistakes form the evidence that can be used to support the assessor judgment during any follow-up dispute that may arise. To gather such evidence, assessors photograph trainee mistakes, enabling the assessor to factually demonstrate the trainee failure and support the original assessment judgement. In addition to using this archived data was during disputes, assessors also use it to familiarise themselves with a trainee who re-attempting an assessment.

3.3 Concerns Theme

During the thematic analysis concerns formed the final theme identified from the assessors interviews, this theme was formed by the following sub-themes, see figure 3. Experts expressed concerns regarding *unsatisfactory automatic feedback* having doubts about the the feasibility of automatic feedback and assessment provided by a computer system. Moreover, concerns were expressed with the viability of the technology to provide an authentic training experience similar to that experienced in real-world training, such as lack of valid tactile feedback for operating equipment, along with concerns regarding their overall lack of familiarly with VR technology. Showing that assessors were *hesitant about technology* adoption in their practice. assessors also pointed out that digital simulations or learning would be prone to *abuse or cheating* by the trainees who may exploit bugs, glitches or limitations in the simulation. *Lack of instant communication* when reviewing feedback, would be problematic for the experts, especially if they could not intervene for mistakes instantly. Finally, concerns about the role of the assessor would be reduced to being *passive observer* in the training process, with no useful input on the learning and creating situations where they would be prone to getting distract when passively watching the trainee performance.

4 Functional Requirements

The themes presented in section 3.2 can be used to define the requirements for tools that enable authentic assessment in room-scale VR, ensuring that such tools satisfy the needs of assessors to adapt classroom-based training to VR training, conduct authentic assessment, and generate contextual feedback.

The aim of authentic assessment is to observe and judge the trainee against a set of criteria to determine the trainee competency [64]. To achieve this tools will have to:

R1 provide direct and unrestricted control of the point of view in the trainee performance without hindering the trainees ability to operate the training simulation and complete the tasks.
R2 enable the assessor to make such judgement by monitoring and observing the trainee applying their knowledge to the tasked objectives.

By providing assisted assessment aids that streamline the assessment for the assessor, the capabilities and opportunities of the assessor should be enhanced, through improved observation and assessment capabilities. The assessor should be able to:

R3 quickly identify the task the system logged as passed and failures.
R4 identify key-points of the trainee performance to speed up or skip down-time in the simulation when the trainee is reading instructions and yet to start any objectives.
R5 identify the confidence of the trainee, so assessors can provide additional supportive feedback for trainees that appear to be under or over-confident.

Feedback enables trainees to reflect on their experience with the input from an assessor used to improve trainee confidence, and correct mistakes or gaps in knowledge. In classroom-based training both the assessor and trainee would be present within the same context, however with virtual feedback being conducted as after-action and detached, trainee and assessors would be in different contexts with the risk that the nuances of physical gestures from real-world feedback would be lost. To mitigate this risk, feedback must be able to:

R6 generate contextual evidence for the trainee that presents a clear indication of the feedback for when, where and why it is being generated.
R7 a record of trainee performances for be kept to be used either as evidence for disputed results or to monitor trainees who are reattempting assessments.
R8 supplement for the lack of gestures enabling the assessor to clearly and concisely focus the attention of the trainee on exact feedback points.

5 Designing and developing the tools

5.1 Tools Identification

To address the functional requirements identified in section 4 the following tools were identified as viable solutions.

As assessors should be able to observe the training simulation from any perspective, a *free-camera tool* was identified as a good solution to address this problem, this was influenced by the free-camera used in existing implementations of virtual performance observation [58, 52]. The free-camera tool should allow assessors to alter point of view dynamically, monitoring the trainee performance from any position or angle, removing the constraints of assessment through video recordings, which are restricted to a single perspective [59]. This ability means that nothing is hidden or obscured from the assessor view of the trainee performance, since they have complete control of their observation.

Like many existing implementations of observational review features [58, 52, 59], *replay features* including the ability to pause, rewind and skip ahead should be available to assessors.

The combination of the free-camera tools and replay features aim to address requirements R1, R2 and partially R4, R5 and R6 in section 4.

Summative performance indicators of the learning outcomes automatically generated by the simulation should be presented to the assessor to provide a general overview of the trainee performance for pass or failure of objectives of the trainee to familiarise and prepare them for assessment.

An *augmented timeline* was identified as a way to show key moments in the trainee performance. The augmented timeline should show information about the trainee performance against time, time-stamping when data modification occurs, thus creating an easy to digest method of identifying key-points during the trainee performance.

Tools should be made available to *assist the visualisation of the trainee performance* in a manner that best represents the trainee actions for visual clarity, which should include visualisation of the trainee posture, providing a simple indication of their stance and movement while completing actions.

The summative performance indicators combined with the augmented timeline and the assisted the visualisation of the trainee performance aim to address the assessment requirements R3, R4 and R5, and partially R1 and R2 in section 4.

A *video narration tool* should enable assessors to capture clips of the trainee performance in the virtual environment, with the ability for assessors to record audio

discussing and narrating the feedback of the training performance. The final output of narration should include both visual and verbal elements that provide the feedback to the trainee with context to the trainee visual performance and failure or area that feedback references.
An *evidence capture tool* should replicate real-world approaches of storing evidence, enabling assessors to capture an image of the training performance of the trainee, clearly displaying errors made during the trainee performance.
In line with the literature in education [60] assessor highlighted the use of gestures to focus the trainee attention, a *drawing tool* that is compatible with all other feedback tools identified should be generated as replacement for the gestures that assessor would use to direct the trainee attention, allowing assessors to annotate and draw on the recorded three-dimensional environment.
The video narration tool with the ability to capture evidence and make annotations or drawing on the three-dimensional environment where the trainee performance takes place aim to address feedback requirements R6to R8 in section 4.

5.2 Combining the Tools in a Framework

The tools identified in section 5.1 have been streamlined into a framework designed to record data from trainees' VR simulation experiences. The framework is based on environment recording, section 2.2, and operates by recording the modification of virtual data embedded into a simulation experience similar to work by Howie and Gilardi [23], Raij and Lok [52], Stone et al. [58] and Backlund et al. [2].

The difference between the framework and existing approaches such as Howie and Gilardi [23] lies in our patent-pending dual-recording methodology [25] that it is used to capture all the data in the simulation without impacting the simulation run-time. Current approaches to recording have several limiting factors which impact the run-time simulation or do not capture sufficient data [2, 37] for the purposes of authentic assessment. For instance, linear data recording [61] is based on motion capture techniques that record only the trainee data, but not the additional environmental data that is required for authentic assessment [2, 37]. Another example is the impact of recording on the frame-rate [46] using reflection techniques as discussed in [23].
Existing methods for capturing all data within a VR simulation has significant impact on the performance of the simulation, causing 'system-lag' and inducing motion sickness [28]. To overcome these limitations, a novel dual-recording methodology [25] for capturing and reconstructing data in room-scale VR simulations was developed for this framework, and underpins the viability of the framework components. The dual-recording methodology [25] requires developers to partition data in non-deterministic and deterministic during the implementation of the simulation, such data are respectively recorded by the lightweight recording layer and processed recording layer, see Figure 4. When combined, data from the lightweight recording and processed recording form the entirety of the data that represent the trainee performance and the simulation. Non-deterministic data are captured by the lightweight recording layer at run-time, and include user inputs, and AI and physics components that might be affected by numeric errors or do not behave consistently between different simulation runs. Deterministic data are captured by the processed recording layer after the trainee performance ended using the lightweight recording as input in 'off-line' run of the simulation. Deterministic data consists of all remaining elements of the simulation that given the same input produce exactly the same data, for instance attaching an object to a controller when the input from the user that the grab button is down is detected. This dual-approach overcomes difficulties encountered recording methods that try to record all

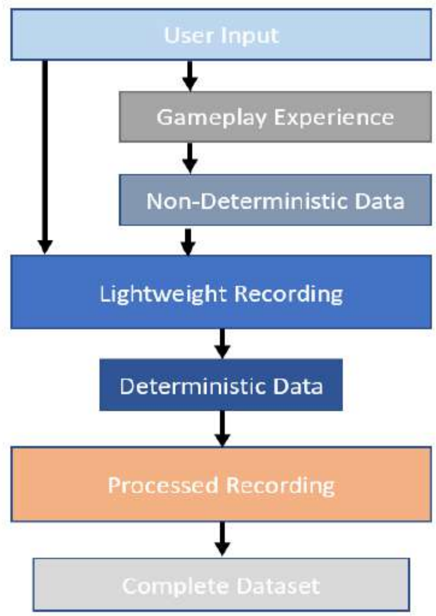


Fig. 4: Overview of the data accessed by lightweight recording and processed recording. The output generates a complete dataset of all simulation data.

information in one go, which can potentially create out-of-sync data [9]. The dual recording approach allows to record all simulation data, enabling adaptation to suit the needs of assessors during the assessment, without the restriction imposed by methods that needs to know in advance which data must be recorded in the simulation, such as [52], [43], [29], and [8]. This data is stored within a database that allows to query information and reconstruct the simulation, as discussed in section 5.3. A framework overview is given in figure 5, which shows that, in addition to a database, the framework is composed by five main components.: lightweight recording, processed recording, observation, assessment and feedback. Lightweight and processed recording form the dual-recording methodology [25], which captures all data of a VR trainee performance within the simulation. Such data is then used to reconstruct and reproduce the simulation enabling assessors to use the observation (section 5.4), assessment (section 5.5), and feedback (section 5.6) components of the framework to do authentic assessment on the trainee performance. These components will be briefly discussed in the reminder of this paper.

5.3 The Database

The flexibility of the dual recording methodology allows developers and assessors to avoid explicitly defining observation or assessment variables that are external to the simulation logic. As such, common assessment and performance values, such as, but not limited to, time completion, number of errors, motion data, objectives completed [34, 33, 43] are accessible by querying the database, see Figure 6, and require no prior configuration by the developer, as is demonstrated by the summative performance indicator interface of the framework, see section 5.5.1. Within the framework presented, the database acts as a source of information that provides the ability to implement the assisted assessment and feedback tools and features that make the assessors job easier [21]. As a consequence of having the database, the observation process that is part of the assessment experience goes beyond visual reproduction of data.

The database stores and manages all data within the framework. The database mimics the structure and hierarchy of the data used by the game engine (in our implementation it was the Unity game engine but any other engine can be used). To make information of the trainee performance accessible queries and common search commands are embedded into the framework, for instance by referencing the unique object identifier, component type, and variable name, any variable that has been captured during recording and is stored in the database can be quickly retrieved. This enables the framework to display the history and all modifications that happened in the simulation as a timeline interface, described in section 5.5.2, provide summative performance information, described in section 5.5.1, and keep the simulation in sync during non-linear reconstruction. Figure 6 shows how the database filters and searches saved data using the framework queries information regarding a variable. By identifying the object, the database then continues through the hierarchy to identify the object components and locate the object variables. Once the data is retrieved, commands that provide quick access to common elements of the data can be added.

By replicating the data hierarchy of the game engine used to implement the simulation within the database, the framework can scale and adapt to any type of simulation built with that engine. Because there no imposed restrictions in capturing only predefined variables [53], the database structure allows assessors with the opportunity to analyse data that initially may not have been considered important during the conversations with developers in the design stage of the simulation [53].

5.4 Observation

For assessors to make use of the data captured from the simulation, an observation component is required within the framework. The observation component reconstructs data saved in the database to reproduce the trainee performance within the simulation using non-linear reconstruction, which allows direct observation of the trainee performance as well as rewinding and time-skipping section of the performance.

During reconstruction the user input data from the database replicates recorded the user actions and objects modifications, enabling the simulation to reproduce the simulation as if a trainee is currently operating the controls. Recorded user input are used to reproduce the trainee movements and actions during their training performance, whilst deterministic and non-deterministic data are reproduced in sync to create a consistent reconstruction of the trainee performance environment modifications.

While the reconstruction of data reproduces a visual trainee performance, the default observation perspective is set to that of the trainee VR view, as in [24, 23]. As discussed in section 2 this view may induce motion-sickness in the observer. To avoid this, the framework also includes a free-camera tool that enables assessors to view the reconstruction from any perspective as dis-

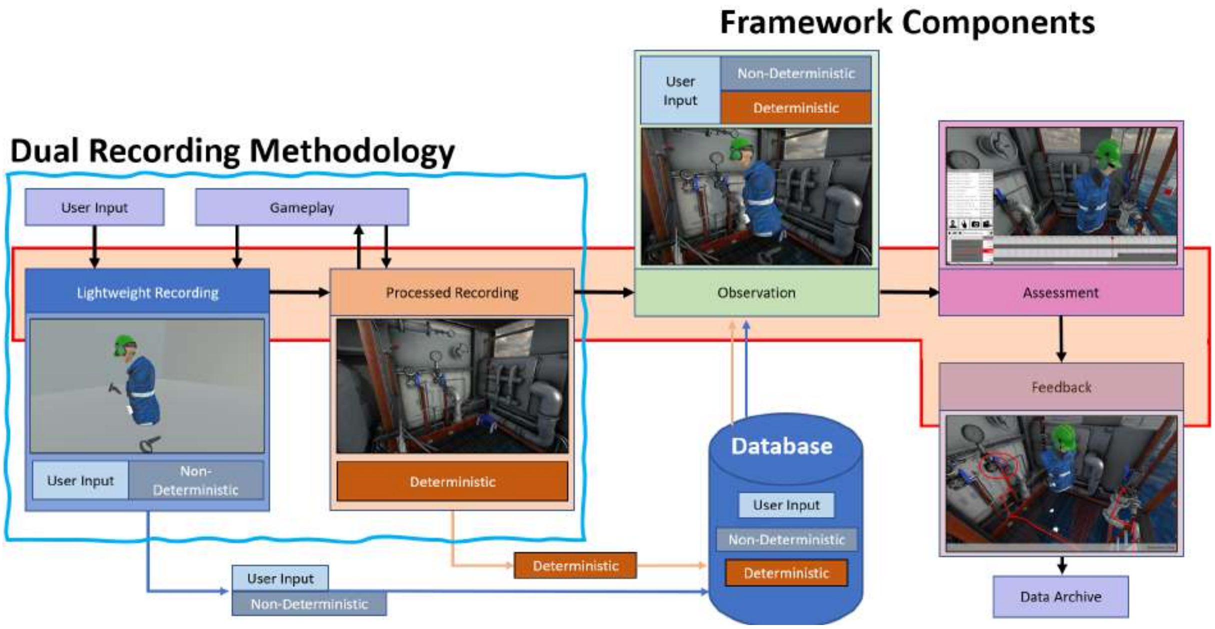


Fig. 5: Overview of the framework components for the dual recording methodology and observation, assessment and feedback operations for conducting assessment and feedback of trainee VR performances.

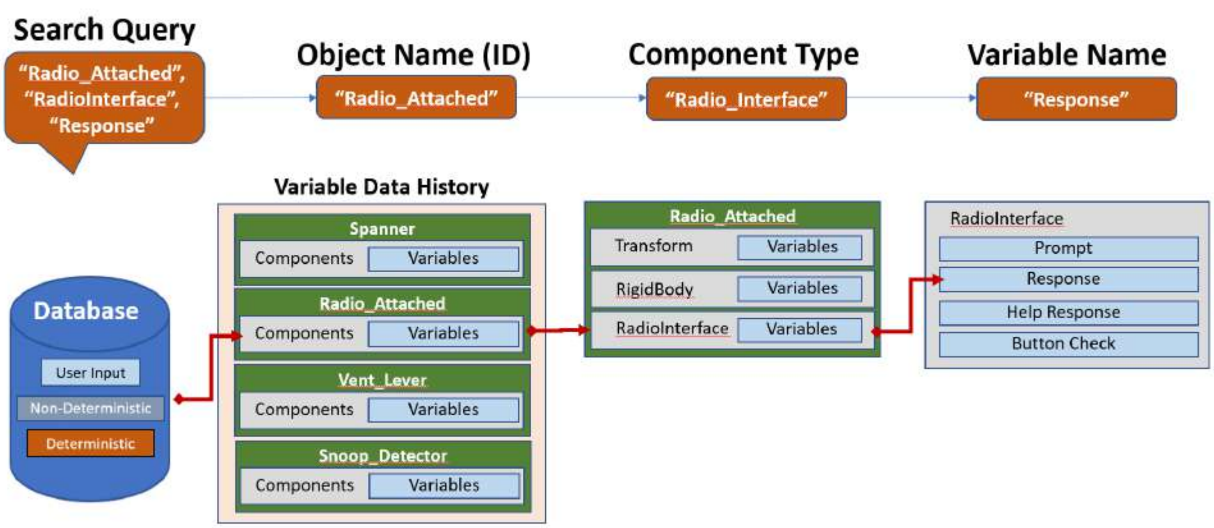


Fig. 6: A search query for the database which looks for the object Radio Attached, in component RadioInterface and variable name Response.

cussed in section 5.1, meeting requirements R1 and R2 in section 4, and partially fulfilling R4, R5, R6.

As the virtual environment is not restricted by the physical constraints of reality, the Observation component is able to provide the assessor with control of the observation, avoiding all concerns of intrusiveness or obscured observations highlighted in section 3.3. The free-camera control features was designed to replicate the conventional approach of exploring virtual worlds using a free-camera system, which operated using keyboard controls to alter the position of the camera, with mouse movement diving the rotation of the camera. When combined, the controls emulate the assessor moving to gain a better view of the trainees actions in the classroom, providing the assessor with tools to observe the performance of a trainee clearly.

With all variable data from the virtual training performance captured, the observation component is able to use non-linear reconstruction. Unlike linear reconstruction methodologies [61, 24] which must be played in order from the very start of the data-file, non-linear reconstruction enables jumping to different points in time, ensuring consistent reconstruction. The observation component includes additional tools and features that enable rewind, forward, time skipping and action identification.

One of the requirements identified in section 4 was the ability to skip sections of data when the trainee is not performing actions within the simulation, requirement R4. The non-linear reconstruction tool contributes to address R4 enabling assessors to reproduce the trainee performance from any time period and skip irrelevant or uneventful sections of the trainee performance, thus improving time efficiency of the observation.

While these tools and features enable observation of the trainee performance, further assessment and feedback tools are required to allow assessors to conduct assessment and generate feedback.

5.5 Assessment

As discussed in section 5.1 a set of tools that meet requirements R3, R4 and R5 from section 4 was identified. These tools assist authentic assessment activities and include summative performance indicators, and augmented timeline interface. The user interface was created for the assessment functionality of the framework, and was inspired by video editing suites, such as Vegas Pro [39]. The user interface was designed to provide sufficient information to the assessor about the trainee performance for authentic assessment, and provides access to the assisted assessment tools.

5.5.1 Summative performance indicators

Most training simulations have pre-existing objectives that trainee need to accomplish, which need to be monitored to ensure that the trainee is achieving the learning outcomes [8, 43]. Thanks to the dual-recording system described in section 5.2 information about these objectives is implicitly recorded in the database, providing information about whether objective was successfully completed.

Figure 7 shows recorded summative performance indicators for the simulation we used to evaluate the framework. assessors can determine whether the simulation objectives were achieved based on the outcomes of the trainee task status. The data was dynamically gathered during the reconstruction for the observation and assessment components, and added to the interface by looking for the variable names that were associated with the task performance of the trainee. The retrieved value of the variables is then compared against an expectation predefined at design time to determine if an objective in the training simulation has been completed. The data is then displayed on the summative performance indicators interface with passed or failed conditions, with timestamps of the relevant variable value.

Summative performance indicators were used to enable assessors to quickly familiarise themselves with the trainee performance prior to the observation. The summative performance information can be provided as a reference to identify when and where a trainee has failed an objective in the simulation, allowing the assessor to pay attention to these areas of the trainees performance, while providing an overall overview of the objectives completed. The summative performance indicator fulfills requirement R3 and contributes to the fulfillment of requirement R4 identified in section 4.

5.5.2 Augmented Timeline

To enable assessors to identify key-points of trainee performance for requirement R4, and partially R1 and R2 in section 4 is the augmented timeline interface. In real-world assessment, familiarisation makes identification of trainee progress easy to acquire, allowing experts to focus attention on areas of concern where failure is prevalent. However, within VR experiences, it is difficult to adapt that real-world identification of progress to suit changes in assessment approach. This can be partially explained due to the change in context, with the virtual environment replicating a real-world scenario, rather than a classroom-based training configuration. To provide an comprehensive overview of the trainee performance, a timeline was implemented to assist assessors in determining to clarify the actions of the trainee against the time of their performance. Like for the summative performance information, by displaying the history of any variable recorded in the database, the variables presented on the timeline can be altered to suit the needs of the assessment.

When starting an observation or assessment of trainee performance, the timeline interface queries the database of recorded trainee data to access the variable history from a predefined list. The retrieved values are then populated on the timeline, creating a new row for each variable, thus generating the variable history for each timestamped modification. The timeline interface includes three key areas, the variable name, current value of the variable and the complete history of variable information. This history is scaled to blocks of data that represent the value of the variable. If variable from the database has no modification history, the timeline marks the row as red to alert to the assessor that objects that variable is attached to were not modified during the simulation, see Figure 7. The timeline can be analysed by the assessor observing the trainee performance, using the data as guidance to jump and observe relevant points of the simulation. For the example in Figure 7, the assessor can use the timeline to identify exactly when in the simulation the training operator picks up equipment related to an objective, then skip to that point in the simulation while viewing the variable values of other equipment at that timeline instance.

5.5.3 Assisted Visualisation of the Trainee Tools

The final assisted assessment aid that contributes to the fulfillment of requirement R5, and partially R1 and R2 in section 4 is the inclusion of assisted visualisation of the trainee tools, aimed to enable assessors to estimate confidence from trainee performances to identify over or under-confident trainees that may require additional support.

As assessors suggested that posture and body movement was a key sign of estimating the confidence of the trainee. As such, assisted visualisation of the trainee tools include a skeleton structure visualisation tool that provides assessors with the ability to estimate the posture of the trainee when completing tasks and objectives. Assisted visualisation of the trainee tools also automatically captured and reproduced the audio from the trainees microphone from the VR equipment. When the simulation is reproduced for assessment, the audio recording of the trainee is replayed in-sync with all other elements of the simulation.

These tools and features enables assessors to hear and see the posture and movement trainee, providing additional audio and visual cues that can be used to estimate confidence levels of trainees, replicating how confidence was measured in classroom-based assessment (see Section 3.1).

5.6 Feedback

As discussed in section 4 three tools were identified from the requirements R6, R7 and R8, to assist the feedback generation by the assessors: *Video Narration Tool*, *Evidence Capture Tool* and *Drawing Tool*.

5.6.1 Video Narration Tool

As discussed in section 3.2.2, the feedback features theme highlights the importance of providing feedback that enables trainees to identify and understand what went wrong during the examination and correct their gaps in knowledge. As discussed in section 4 a video narration tool was identified as best suited to provide contextual and visual evidence with the feedback.

The video narration tool captures video recordings from the perspective of the assessor's free-camera when observing the trainee performance, capturing exactly what the assessor is looking at, including any movement or changes in perspective from the camera. Audio recording from the assessors microphone are also captured and embedded into the generated video, enabling narration of the feedback within the context of the trainee performance. The technical aspects for recording the video was handled by incorporating in the framework a third party plug-in Video Capture Pro [10] that enabled the video and audio recordings captured by the assessor to be exported as .mp4 and accessed outside of the Untiy engine environment [10]. The video narration tool allows to generate video evidence of the trainee performance as the videos generated by the tool can be archived, partially fulfilling requirement R7 and fulfilling requirement R6 from section 4.

5.6.2 Evidence Capture Tool

As discussed in section 3.2.2 evidence need to be generated and archived when trainees are likely to fail an assessment. In section 5.1 an evidence capture tool was identified to fulfil this need, requirement R7 from section 4. Such tool would allow assessors to provide evidence when trainees disputed their outcome. While disputing assessment outcomes is considered rare by assessors, it represents a problem that may be more common in virtual forms of assessment as concerns were risen about cheating or abuse of the system in section 3.3. Moreover, trainees may feel inclined to dispute any assessment of their performance on the assumption the assessor would be unable to dispute their claims. While the framework enables the entire trainee performance data-set to be archived, or to video record the performance using video narration tool, capturing a picture of a failure mimics current practice adopted by assessors during assessment. The evidence capture tool captures a single image of the trainee session from the assessors perspective, outputting an image, as shown in Figure 8.

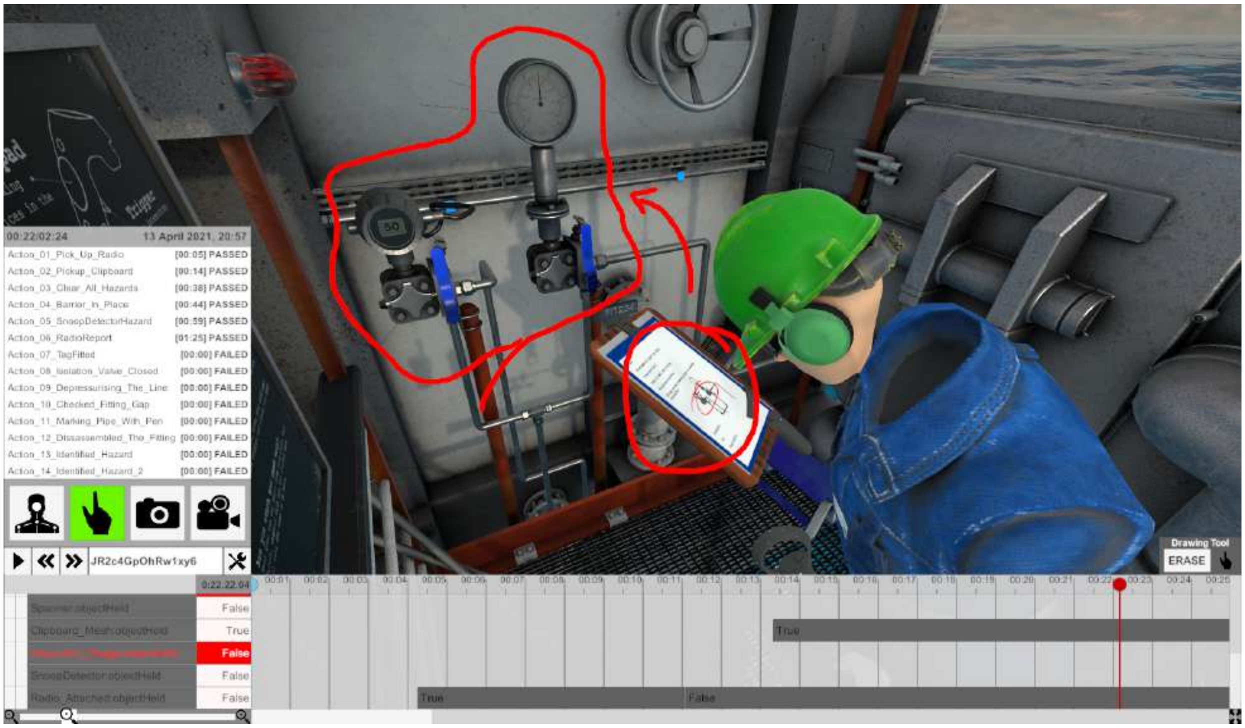


Fig. 7: The interface implementing tools identified in section 4, such as the summative performance indicators, the timeline, the play, pause, rewind, and forward buttons, along with other elements of the interface including the drawing tool and video narration tool.

5.6.3 Drawing Tool

Gestures used by assessors to direct the trainee attention during feedback were identified as one of the key practices adopted by the assessors, see section 3.2.2. During feedback using VR environments, pointing and gesturing to trainee is not possible. To ground verbal and visual feedback to be correctly to a specific area of a simulation a drawing tool was identified in section 5.1 and implemented into the framework. Such tool allows the assessor to 'draw on top of' three-dimensional space from the point of view of the free-camera, enabling lines and drawn text to be embedded from the assessor perspective in both video and pictures captured using the tools discussed in this section. While this tool is not a like-for-like replacement for real-world gestures, it is an accessible approximation that assessors can use to focus the attention of the trainee during feedback to highlight objectives that were missed by the trainee.

The drawing tool can be seen in use in Figure 7, whilst Figure 8 shows the output generated by the video narration tool combined with the drawing tool to create contextual and targeted feedback for the trainee. This tool fulfills requirement R8, and partially R6 from section 4.

6 Evaluation

To validate the framework, an implementation of it was developed using the Unity game engine and qualitatively evaluated with assessors within industry and academia. The study was designed as a factorial within subjects non-experimental evaluation, in which factors considered were the framework and two existing authentic assessment methods for VR room-scale simulations, namely analytical data and video recordings. Assessors where observed using all three methods to conduct authentic assessment and generate contextual feedback for two virtual reality training performances, participants were recorded and notes were taken during the observations. The assessment and feedback generation stage was then followed by a semi-structured interview to allow the assessors to express their thoughts, feelings and impression of the three methods. Six assessors from the oil-and-gas training industry and seven assessors from academia were recruited this sample size was deemed acceptable based on Malterud et al. [40] information power theory. Thematic analysis was performed to analyse the data collected during the evaluation following the process described by Braun and Clarke [6]. After familiarization with the transcripts, initial codes were generated and grouped into themes, the themes where then discussed with the second author and finalised after reaching and agreement.

Fig. 8: The output of the video narration tool, showing the simulation time that references the feedback and the overlaid drawn gestures from the drawing tool.

The methodology adopted during the evaluation focused the direction of the evaluation on four key topics:

- Usability of the framework and its features.
- Usefulness of the framework and its features.
- Comparison of the usability of the framework against the other two methods.
- Comparison of the usefulness of the framework against the other two methods.

6.1 Key Findings

From the thematic analysis, four key themes emerged: *Audio and Visual Cues*,*Assisted Assessment and Feedback*, *Perspective of Assessment* and *Follow-up Dialogue Session.*

6.1.1 Audio and Visual Cues

It was identified that without direct observation of the trainee, no context of the trainee action or verbal comments were possible for assessment. With visual and verbal cues, the trainee performance was put in context allowing assessors to identify what the trainee was doing and how they are coping with the tasks. Visual cues provide the context to what the trainee does during the assessment and the audio cues provide information for how the trainee is coping.

Without visual and audio cues from the trainee performance, assessors reported that they did not have confidence in the predefined learning that were automatically generated by the analytical factor, and were unsure on how to approach or provide assessment and feedback, as observed by Participant 7:

> *"Based on what we do just now [in classroom-based training], we physically see them and we observe them. It could be that [he] was just lucky guessing, I just don't know." (Participant 7)*

As authentic assessment focuses on a method of assessing learning which is practical and natural for the real-world conditions [64, 63], visual and audio cues during assessment enabled assessors to see the trainee logically apply their knowledge to resolve the problem. Therefore, without the visual and audio context, the outcomes of the trainee performance were considered to have no merit associated with them, and outcomes from analytical data were not trusted or considered reliable indicators of competency in the eyes of assessors. In comparison, the factors that had audio-visual components of assessment, i.e. video recording and the framework, enabled assessors to understand the motivations and thought-process the trainee, and to identify areas of low confidence in the trainee performance from the verbal comments made by the trainee, as noted by Participant 8:

> *"He was slow, but methodical, and he was sort of questioning himself." (Participant 8)*

When combined, audio and visual cues provide a strong contextual reference to both the actions conducted by the trainee and their thought process when performing actions. Information from these cues provides a reliable platform which presents assessors with in-depth context and knowledge of the trainee performance, enabling assessors to determine if trainees are competently performing the objectives. Assessors almost always referenced a combination of the audio and visual cues when judging the confidence level of the trainee, allowing assessors to understand the context for where, when and why the trainee is struggling in the simulation.

6.1.2 Assisted Assessment and Feedback

The second key finding showed the importance of embedding assisted assessment and feedback tools into a single framework environment for conducting both assessment and feedback, streamlining the process of conducting assessment for assessors.

With prior knowledge of the trainee performance, assessors would skip sections of the video factor and focus on the events where the trainee was completing actions. With knowledge for the simulation outcomes, assessors skipped ahead looking to see when elements of the virtual environment were different. In doing so, assessors would ignore all other information provided by the video, focusing on the modification of these changes to identify what stage the trainee was at and to determine if the task was completed, as Participant 11 highlights:

> *"I'm looking to try and find the key points in the assessment where they would carry out a certain activity [...] or they would carry out a particular procedure. You might jump to the key points. And then if you're unsure, go back and watch that action in more detail." (Participant 11)*

As the framework provided this information through the summative performance indicators and the augmented timeline, the framework was best suited to meeting the needs of assessors, providing the necessary prior knowledge that helped assessors identify and areas of concern on the lead-up to a trainee failure in their performance. Information from the trainee performance overlaid on the observation assisted assessor with a reliable reference to how the trainee has performed in predefined objective tasks. With this information, the observation of the trainee can be focused on areas of interest within the simulation. Furthermore, these findings support requirements R1, R2 and R5, justifying the framework design and implementation, which sought to embed assessment aids into the framework that provide identification of actions and events within the trainee performance.

The framework embedded contextual feedback tools, which included the video narration tool (Section 5.6.1), the evidence capture tool (Section 5.6.2), and the drawing tool (Section 5.6.3). Within the framework the authentic assessment and contextual feedback process was streamlined for assessor, making their job easier [21] and enabling feedback to be targeted to area of relevance for the trainee, as stated by Participant 9:

> *"This is a more comprehensive way of [providing feedback], and ultimately, it is more time efficient." (Participant 9)*

Video recordings required assessment and feedback to be conducted separately. Videos recordings are generated externally to the assessment environment, as such they resulted to be less efficient and effective, impacting the cognitive load for the assessor, as their attention becomes diverted between multiple applications. On providing feedback using video recordings Participant 5 commented:

> *"The process is cumbersome for providing feedback. It's not smooth, and you can't really identify clearly, what went wrong without a lot of jumping through hoops or using [Microsoft] teams, using email, using phone calls. And it's easy to see that the information that you generate for [the trainee] about the failures could get misconstrued or lost." (Participant 5)*

By comparison, the framework do not presents these issues as it combines the assessment and feedback within a single environment. Assessors considered both the process and content of the feedback generated from the framework as very effective, with drawing tool being used by assessors to target and focus their feedback [60].

> *"The amount of detail you can get in the video capture is a brilliant feature [...], the drawing is very, very useful. [...] I would say that ultimately, all of that builds to create in a very comprehensive look at what the trainee does, and where they go wrong. And being able to feed that back in very accurately and detailed [...] it's a very, very comprehensive, and an excellent piece of equipment." (Participant 5)*

As Jonsson [31] note, if feedback is delayed for too long, it becomes redundant, the findings presented are significant, as they show that streamlining assessment and feedback by embedding feedback tools into the same environment as the assessment improves the quality of the feedback provided to trainees.

6.1.3 Point of observation during Assessment

Contrasting themes for the framework and video factors emerged regarding the point of observation during assessment. The framework presented an unrestricted approach for observation of the trainee performance, which enabled assessors to view the trainee performance from any perspective. By contrast, the video factor required assessors to observe the performance from the perspective of the trainee. This point of view was useful as gave assessors the ability to see the training performance through the eyes of the trainee, creating the opportunity to understand the trainees thought process when conducting tasks and objectives, letting them read the situation, and understand the logic and context behind the actions of the trainee, Participant 10 observed:

> *"When you can see the vision through his eyes, I could see what he sees [...] You know, he was looking at lever he was looking at the gauge, he*

was looking at the lever. So I think my confidence [in the trainee] went up a bit when I saw that." (Participant 10)

Using the eye-gaze of the trainee, assessors were able to identify what the trainee was looking in the simulation environment to determine their thought process. By understanding the perspective of the trainee, assessors are guided in their assessment process, improving their ability to relate to the trainee. However, observing the performance from this perspective induced problems for assessors inducing motion sickness or discomfort in 5 of the 13 participants, as evidenced by Participant 4 comment:

"I was just getting a little bit dizzy as [the trainee head] was moving around [...] the movements in the video are very erratic, because you're looking at it from the [trainee] perspective." (Participant 4)

While multiple static perspectives could be used to overcome this issue [59], long term assessment or observation from the perspective of the trainee appears to have implications for the health and well-being of assessors.

However, assessors appreciated the information acquired by viewing performance form the trainee point of view. Participant 13 suggested a variation of it as an improvement to the framework:

"I think with the [framework] an improvement would be so you can also watch it from the trainee's perspective [like the video factor]. If you could do it from both, that would be more valuable. Like if you can use two screens or something, you know, you could have [the trainee's perspective] and at the same time using the [free camera tool], a small picture and picture type thing. That would be really good." (Participant 13)

As such, the use of the trainee perspective acts as an additional assessment aid, rather than being the primary method of observation, as was the case with the video factor. Operating as a picture-in-picture (PiP) tool, the perspective of the trainee can be used to enhance the information available to the trainee. However, this PiP implementation may still cause discomfort or motion sickness for assessors and can potentially caus split attention issues. Alternatively, eye-tracking equipment could be used to mitigate this issue [42], allowing the trainee gaze to be visualised, providing context and information to the assessor as an additional feature within the assisted visualisation of the trainee tools.

6.1.4 Follow-up Dialogue Session

The final key finding from this evaluation was the prevalence of demand for a follow-up session with the trainee which would involve the assessor and trainee reviewing the assessment together to discuss the learning outcomes, and provide the trainee with the opportunity to directly question the assessor. These findings were prevalent in both academic and industry responses, indicating that Q&A sessions are an important element within learning [50], potentially because it forces learners to review the feedback they receive in preparation [31]. As stated by Jonsson [31], if learners do not view or utilise the feedback, they do not learn from their mistakes. Therefore, by following up on the trainee, the assessor can determine if the trainee has understood the feedback, and check if they need any further help or support. While Hanoun and Nahavandi [21] argues that performance must be reviewed and acted upon immediately after an assessment, this is not always possible for remote forms of assessment. Time-zone differences or commercial considerations create situations where there is a delay between the completion of an assessment by a trainee, and the feedback received by a assessor. Therefore, for authentic assessment, it is best to focus on the process of streamlining the assessment and generation of feedback, which can then be followed-up with a dialogue session.

As participants were not familiar with VR, there was a concern that when generating feedback remotely, trainee may not understand the feedback. To mitigate this concern, assessors considered the follow-up dialogue session as important to allowing them to check on the trainee and see how they are coping, and determine if any further assistance is required. Because it is difficult to determine the trainee response to feedback remotely, follow-up dialogue sessions were seen as an opportunity to enable one-on-one discussion with trainees. However, comments related to follow up dialogue sessions appeared predominantly for the analytical data and the video recording factors, where the assessors felt the generated feedback was poor or insufficient for correcting major shortcomings in a trainee performance. The feedback supplied with the framework was considered equivalent to that provided in traditional classroom-based environments, which included context and targeted feedback, and there was less demand for follow-up dialogue sessions. When asking how the feedback compared to classroom-based feedback Participant 9 commented:

"Pretty amazing, really, if that's a way of giving feedback over what we did previously, which was just, you know, me talking to maybe send

> *them an email or something, it seems quite archaic and I get this kind of system. It's pretty amazing way of giving feedback. And if you're talking about doing it on the other side of the world, you're losing that face to face thing, but I think you're able to explain things so much more clearly." (Participant 9)*

However, dialogue was still considered a useful education and support tool for instances where trainees needed further prompting.

6.2 Discussion of Framework Tools

The framework included several tools and features which were evaluated by assessors who discussed their use and influence on the assessment process.

6.2.1 Observation

To provide direct visual observation of the trainee, the framework embedded a *free camera tool* and *replay features*, which enabled assessors to view the trainee performance from any perspective at any time period.

Free camera tool Findings from the study in section 6.1 show that the free camera tool was effective for assessors, allowing observation without any restrictions, obstructions or concerns that their gaze would impact the trainee. As such, the free camera tool (see Section 5.4) met the requirements R1, R2, R5 and R6 from Section 4) and successfully completed the other observation, assessment and feedback tools and features.

Replay features The replay features met the requirements of R1, R2 and R4 from the background study conducted with assessors. The replay features provided the ability for assessment to be replayed, or skipped as required (see Section 5.4), overcoming concerns discovered with classroom-based assessment.

6.2.2 Assessment Tools

Assessment tools included the *summative performance indicators*, *augmented timeline* and the *assisted visualisation of the trainee tools*, which provided enhanced abilities that improved the assessment process for assessors.

Summative Performance Indicators The summative performance indicators provided assessors with the means of quickly familiarising themselves with the trainee performance and were seen as an additional assessment aid that provided assessors with knowledge of the trainee performance prior to beginning the assessment. However, assessors stressed that alone the summative performance indicators would be inadequate, and required context to through direct visual observation to conduct assessment and generate feedback. These findings show that the summative performance indicators (see Section 5.5.1) successfully met the requirement R3 from Section 4 for assessors.

Augmented Timeline The augmented timeline was found to improve the assessment process, with assessors using the additional information to influence their assessment of the trainee performance, enabling deeper insight into the performance of the trainee. The augmented timeline provided in-depth context to the actions of a trainee performance, displaying key-points of trainee data in-sync with the direct visual observation of the trainee, allowing easy identification for the progress of the trainee. These findings show assessors used the augmented timeline to improve the process of conducting assessment. As such, the augmented timeline (see Section 5.5.2) met assessors requirements of R1, R2 and R4 which were identified in Section 4.

Assisted Visualisation of the Trainee Tools Assisted visualisation of the trainee tools included a skeleton structure visualisation to identify the posture and movement of the trainee, along with the audio cues captured from the microphone of the trainee. As already discussed in Section 6.1.1, audio cues were a key finding of this evaluation, signifying the importance of audio cues in providing additional information that can contextualise the trainee performance for the assessor. A skeleton structure visualisation was added to these tools, but assessors considered it unnecessary because the default virtual avatar of the trainee provided a sufficient representation of the trainee. However, assessors were observed using the skeleton structure visualisation to hide the default virtual avatar and have a better view of what the trainee was doing. The irrelevance of the posture for the assessors was probably caused by the specific evaluation scenario used in this study and other training simulation scenarios may require actions and objectives to be completed using specific postures [4]. These findings indicate that the assisted visualisation of the trainee tools met the requirements R1, R2 and R5 from Section 4).

6.2.3 Feedback Tools

Feedback tools include the *video narration tool*, *evidence capture tool* and *drawing tool*, which when combined with the assessment tools (see Section 6.2.2), streamlined the assessment process for the assessor.

Video Narration Tool The video narration tool was often viewed as the best method of providing contextual feedback in virtual training simulations, with many assessors praising its implementation, along with the feedback it generated. Assessors reflected that the feedback supplied using this tool was far beyond their expectations, referring to existing methods of feedback as inadequate. These findings imply that the video tool satisfies requirements R6 and R7 from the background study conducted with assessors in Section 4, with design of the tool (see Section 5.6.1) proving successful to achieve comprehensive feedback that is comparable with classroom-based feedback.

Evidence Capture Tool The evidence capture tool was restricted to capturing only an image of the trainee performance to match classroom-based usage, fulfilling requirement R7 in Section 4. However, assessors considered the evidence capture tool as a step back, preferring the video narration tool to gather evidence.

Drawing Tool The drawing tool was considered an important aspect of the feedback process, with assessors discussing its positive impact on their ability to focus their feedback. The drawing tool enabled the feedback generated from video narration tool to be more effective and comprehensive. These findings show that the drawing tool was considered as an effective method that further extended the capabilities of the assessor and was a suitable replacement for gestures, meeting requirements of R6 and R8 from Section 4, enabling feedback to be richer in detail, improving its overall quality and usefulness.

In summary, 11 out of 13 assessors ranked the framework as their first-choice for conducting assessment and feedback for room-scale VR training simulations. Findings conclude that the overwhelming preference of the framework as the preferred method was due to the tools provided. Because these tools were embedded into the framework, assessors praised the ability to quickly and efficiently generate feedback, without diverting their attention to other applications or losing track of their assessment progress. Furthermore, the assisted assessment tools and features enabled in-depth assessment of trainee performances, which allowed assessors to engage with the data to analyse the augmented timeline and summative performance information in sync with their visual observation. These findings justify the approach used to conduct the research, highlighting that the existing research which influenced the design of the framework components [21, 58, 52], has been validated in practice.

7 Conclusion

In this paper tools for authentic assessment and contextual feedback have been identified from a contextual user study with assessors from industry. The tools design and their structuring into a streamlined framework was then discussed. A novel dual-recording methodology [25] was briefly described in this paper and forms the foundation of the framework, enabling the capturing of the data required by the tools described. Finally, an implementation of the framework was created to evaluate the usefulness of such tools in academia and industry. The evaluation highlighted that the tools presented in this paper improved the usefulness and usability of authentic assessment and contextual feedback for room-scale VR training. The augmented timeline and the summative performance information were useful to identify the areas of immediate concern, lessening the cognitive overhead for assessors. These assisted features could be further enhanced with transcribed audio logs of the trainee [52] to ensure that context remains for sections that are skipped by the assessor. Embedded contextual feedback tools demonstrated effective generation of targeted feedback that assessors considered useful and usable to support the learning of trainees. The combination of the assessment and feedback tools in a single framework streamlined authentic assessment operations, simplifying the process and lessening the cognitive overhead for the assessor. The need for a follow-up dialogue session was noted to be dependant on the performance of the trainee and the methods used to generate feedback, i.e analytic data and video recordings. For the framework the follow-up dialogue session was considered as a useful optional step for the feedback process. Although results from the evaluation are encouraging further work is required to improve the interaction with the camera, as some assessors were unfamiliar with methods used in computer games to control a free camera using keyboard and mouse, requiring them to spend some time to familiarise themselves with the controls. In addition, further research will be conducted into implementing AI tools that assist in analysing and displaying data recorded from the dual-recording methodology. Remote functionalities as described in [23] will be also integrated and evaluated in our future efforts.

Acknowledgements We would like to thank Swagelok Scotland ltd for permission to adapt and re-use their commercial VR training simulation integrating the new tools and framework into it. We would also thank their staff that participated in the requirement collection and evaluation. And and the University staff that participated in the evaluation study.

References

1. de Armas C, Tori R, Netto AV (2020) Use of virtual reality simulators for training programs in the areas

of security and defense: a systematic review. Multimedia Tools and Applications 79(5):3495–3515

2. Backlund P, Engstrom H, Hammar C, Johannesson M, Lebram M (2007) Sidh–a game based firefighter training simulation. In: 2007 11th International Conference Information Visualization (IV'07), IEEE, pp 899–907
3. Beller M (2013) Technologies in Large-Scale Assessments: New Directions, Challenges, and Opportunities, Springer Netherlands, Dordrecht, pp 25–45. DOI 10.1007/978-94-007-4629-9_3, URL https://doi.org/10.1007/978-94-007-4629-9_3
4. Bernard F, Zare M, Sagot JC, Paquin R (2018) Virtual reality simulation and ergonomics assessment in aviation maintainability. In: Congress of the International Ergonomics Association, Springer, pp 141–154
5. Blomberg J, Burrell M, Guest G (2007) The human-computer interaction handbook. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, chap An Ethnographic Approach to Design, pp 964–986, URL http://dl.acm.org/citation.cfm?id=772072.772133
6. Braun V, Clarke V (2006) Using thematic analysis in psychology. Qualitative Research in Psychology 3(2):77–101, DOI 10.1191/1478088706qp063oa, URL https://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa, https://www.tandfonline.com/doi/pdf/10.1191/1478088706qp063oa
7. Bric JD, Lumbard DC, Frelich MJ, Gould JC (2016) Current state of virtual reality simulation in robotic surgery training: a review. Surgical endoscopy 30(6):2169–2178
8. Carruth DW (2017) Virtual reality for education and workforce training. In: 2017 15th International Conference on Emerging eLearning Technologies and Applications (ICETA), IEEE, pp 1–6
9. Epic Games (2021) Replay system. URL https://docs.unrealengine.com/en-US/Engine/Replay/Streamers/index.html, (Accessed on 16/01/2021)
10. EveReal (2021) Video capture pro. URL https://assetstore.unity.com/packages/tools/video/video-capture-pro-155663, (Accessed on 16/04/2021)
11. Fardinpour A, Dreher H (2012) Towards developing automated assessment in virtual learning environments. In: IADIS International Conference on Internet Technologies & Society
12. FitzGerald E (2012) Analysing video and audio data: Existing approaches and new innovations. Open Reserach Online
13. Fukkink RG, Trienekens N, Kramer LJ (2011) Video feedback in education and training: Putting learning in the picture. Educational Psychology Review 23(1):45–63
14. Goderstad J, Sandvik L, Fosse E, Lieng M (2019) Development and validation of a general and easy assessable scoring system for laparoscopic skills using a virtual reality simulator. European Journal of Obstetrics & Gynecology and Reproductive Biology: X 4:100092
15. Goldberg SL, Knerr BW, Grosse J (2003) Training dismounted combatants in virtual environments. Tech. rep., Army Research Development and Engineering Command Orlando FL Simulation ...
16. Gould D, Kelly D, Goldstone L, Gammon J (2001) Examining the validity of pressure ulcer risk assessment scales: developing and using illustrated patient simulations to collect the data information point: Visual analogue scale. Journal of clinical nursing 10(5):697–706
17. Greenhalgh C, Flintham M, Purbrick J, Benford S (2002) Applications of temporal links: Recording and replaying virtual environments. In: Proceedings IEEE Virtual Reality 2002, IEEE, pp 101–108
18. Grübel J, Weibel R, Jiang MH, Hölscher C, Hackman DA, Schinazi VR (2017) Eve: A framework for experiments in virtual environments. In: Barkowsky T, Burte H, Hölscher C, Schultheis H (eds) Spatial Cognition X, Springer International Publishing, Cham, pp 159–176
19. Gulikers J, Bastiaens T, Kirschner P (2006) Authentic assessment, student and teacher perceptions: the practical value of the five-dimensional framework. Journal of Vocational Education and Training 58(3):337–357
20. Gulikers JT, Bastiaens TJ, Kirschner PA (2004) A five-dimensional framework for authentic assessment. Educational technology research and development 52(3):67
21. Hanoun S, Nahavandi S (2018) Current and future methodologies of after action review in simulation-based training. In: Proceeding of the 2018 Annual IEEE International Systems Conference (SysCon), IEEE, pp 1–6
22. Hattie J, Clarke S (2018) Visible learning: feedback. Routledge
23. Howie S, Gilardi M (2020) Virtual observations: a software tool for contextual observation and assessment of user's actions in virtual reality. Virtual Reality pp 1–14
24. Howie SR, Gilardi M (2019) Virtual observations of virtual reality simulations. In: Proceedings of the 2019 CHI Conference Extended Abstracts on Hu-

man Factors in Computing Systems (To Appear), ACM, New York, NY, USA, CHI EA '19
25. Howie SR, Gilardi M (2021) Methods and systems for recording a user experience
26. Hulsman RL, van der Vloodt J (2015) Self-evaluation and peer-feedback of medical students' communication skills using a web-based video annotation system. exploring content and specificity. Patient Education and Counseling 98(3):356–363
27. Jensen K, Bjerrum F, Hansen HJ, Petersen RH, Pedersen JH, Konge L (2017) Using virtual reality simulation to assess competence in video-assisted thoracoscopic surgery (vats) lobectomy. Surgical endoscopy 31(6):2520–2528
28. Jerald J (2016) The VR Book: Human-Centered Design for Virtual Reality. Association for Computing Machinery and Morgan & Claypool, New York, NY, USA
29. Johnson C, Gonzalez AJ (2008) Automated after action review: State-of-the-art review and trends. The Journal of Defense Modeling and Simulation 5(2):108–121
30. Johnson CI, Priest HA (2014) 19 the feedback principle in multimedia learning. The Cambridge handbook of multimedia learning p 449
31. Jonsson A (2013) Facilitating productive use of feedback in higher education. Active learning in higher education 14(1):63–76
32. Joy Cumming J, Maxwell GS (1999) Contextualising authentic assessment. Assessment in education: Principles, policy & practice 6(2):177–194
33. Konakondla S, Fong R, Schirmer CM (2017) Simulation training in neurosurgery: advances in education and practice. Advances in medical education and practice 8:465
34. Lahanas V, Georgiou E, Loukas C (2016) Surgical simulation training systems: box trainers, virtual reality and augmented reality simulators. International Journal of Advanced Robotics and Automation 1(2):1–9
35. Lazar J, Feng JH, Hochheiser H (2017) Research methods in human-computer interaction. Morgan Kaufmann
36. Lee AT (2017) Flight simulation: virtual environments in aviation. Routledge
37. Lopez T, Dumas O, Danieau F, Leroy B, Mollet N, Vial JF (2017) A playback tool for reviewing vr experiences. In: Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology, ACM, p 83
38. MacCormick D, Sangalang A, Rushing J, Jagpal RS, Mirza-Babaei P, Zaman L (2019) Frvrit: A tool for full body virtual reality game evaluation. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, pp 1–6
39. Magix (2021) Video production software — unleash your creativity. URL `https://www.vegascreativesoftware.com/gb/vegas-pro/`, (Accessed on 01/04/2021)
40. Malterud K, Siersma VD, Guassora AD (2016) Sample size in qualitative interview studies: Guided by information power. Qualitative Health Research 26(13):1753–1760, DOI 10.1177/1049732315617444, URL `https://doi.org/10.1177/1049732315617444`, pMID: 26613970, `https://doi.org/10.1177/1049732315617444`
41. Martins LL, Kellermanns FW (2004) A model of business school students' acceptance of a web-based course management system. Academy of Management Learning & Education 3(1):7–26
42. Meißner M, Pfeiffer J, Pfeiffer T, Oppewal H (2017) Combining virtual reality and mobile eye tracking to provide a naturalistic experimental environment for shopper research. Journal of Business Research
43. Moore AG, Hu X, Eubanks JC, Aiyaz AA, McMahan RP (2020) A formative evaluation methodology for vr training simulations. In: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), IEEE, pp 125–132
44. Newmann FM, Archbald DA (1992) The nature of authentic academic achievement. Toward a new science of educational testing and assessment pp 71–83
45. Nyström A, Pålsson Y, Hofsten A, Häggström E (2014) Nursing students' experiences of being video-recorded during examination in a fictive emergency care situation. International journal of nursing practice 20(5):540–548
46. Oculus (2021) Guidelines for vr performance optimization. URL `https://developer.oculus.com/documentation/native/pc/dg-performance-guidelines/`, (Accessed on 01/04/2021)
47. O'Leary M, Scully D, Karakolidis A, Pitsia V (2018) The state-of-the-art in digital technology-based assessment. European Journal of Education 53(2):160–175
48. Pan JJ, Chang J, Yang X, Liang H, Zhang JJ, Qureshi T, Howell R, Hickish T (2015) Virtual reality training and assessment in laparoscopic rectum surgery. The International Journal of Medical Robotics and Computer Assisted Surgery 11(2):194–209
49. Patle DS, Manca D, Nazir S, Sharma S (2019) Operator training simulators in virtual reality environment for process operators: a review. Virtual Reality 23(3):293–311

50. Pedram S, Palmisano S, Skarbez R, Perez P, Farrelly M (2020) Investigating the process of mine rescuers' safety training with immersive virtual reality: A structural equation modelling approach. Computers & Education p 103891
51. Pietquin O, Hastie H, et al. (2013) A survey on metrics for the evaluation of user simulations. Knowledge Engineering Review 28(1):59–73
52. Raij AB, Lok BC (2008) Ipsviz: An after-action review tool for human-virtual human experiences. In: 2008 IEEE Virtual Reality Conference, IEEE, pp 91–98
53. Raymond C, Usherwood S (2013) Assessment in simulations. Journal of Political Science Education 9(2):157–167
54. Roberts PG, Guyver P, Baldwin M, Akhtar K, Alvand A, Price AJ, Rees JL (2017) Validation of the updated arthros simulator: face and construct validity of a passive haptic virtual reality simulator with novel performance metrics. Knee Surgery, Sports Traumatology, Arthroscopy 25(2):616–625
55. Shute V, Ke F, Wang L (2017) Assessment and adaptation in games. In: Instructional techniques to facilitate learning and motivation of serious games, Springer, pp 59–78
56. Shute VJ (2008) Focus on formative feedback. Review of educational research 78(1):153–189
57. Shute VJ (2011) Stealth assessment in computer-based games to support learning. Computer games and instruction 55(2):503–524
58. Stone R, Snell T, Cooke N (2016) An inexpensive underwater mine countermeasures simulator with real-time 3d after action review. Defence technology 12(5):367–379
59. Thurston CL, Martin GA (2011) A generic after action review capability for game-based training. In: International Conference on Engineering Psychology and Cognitive Ergonomics, Springer, pp 628–634
60. Valenzeno L, Alibali MW, Klatzky R (2003) Teachers' gestures facilitate students' learning: A lesson in symmetry. Contemporary Educational Psychology 28(2):187–204
61. Von Spiczak J, Samset E, DiMaio S, Reitmayr G, Schmalstieg D, Burghart C, Kikinis R (2007) Multimodal event streams for virtual reality. In: Multimedia Computing and Networking 2007, International Society for Optics and Photonics, vol 6504, p 65040M
62. Wan Z, Fang Y (2006) The role of information technology in technology-mediated learning: A review of the past for the future. AMCIS 2006 Proceedings p 253
63. Wiggins G (1990) The case for authentic assessment. Practical assessment, research, and evaluation 2(1):2
64. Wood L, Teras H, Reiners T, Gregory S (2013) The role of gamification and game-based learning in authentic assessment within virtual environments. Research and development in higher education: The place of learning and teaching pp 514–523