

Human Performance Assessment in VR and Desktop Simulated Crane Operations: A Multi-Difficulty Task Study

Yestay Anuarbekov Bekzhan Kimadiyev Shynggys Nakypov
Saken Abdizhamilov Almas Yerbolat

20 April 2026

Abstract

Tower crane work is a safety critical trade that involves accurate motor skills, spatial perception, and sound judgment of the load. Traditional training strategies such as on-the-job training and physical simulators are expensive, not standardizable and dangerous to beginners. This project will fulfill this shortcoming by creating a high-fidelity virtual tower crane training simulator, which will allow training of operators in a safe, repeatable, and cost-effective manner.

The simulator was developed in Unity 2022 LTS and used in two modalities: Virtual Reality (VR) and Desktop. The VR version provides an immersive first-person cabin experience with the use of the Meta Quest 2 and the Desktop version extends the visual experience with first-person, third-person and freely moving camera views. Both modalities have identical crane mechanics, task structure and environment, which facilitate controlled cross-modality comparison. The crane model includes four degrees of freedom (DOF) (rail travel, jib rotation, trolley movement, and hoist control) and dynamic physics of payload oscillation.

The participants were on standardized pick-and-drop tasks with three levels of difficulty: easy, medium and hard. Continuous crane-state variables and discrete operational events during each trial were logged automatically, allowing the quantitative performance analysis of each trial on measures such as task completion time, hook path length, path efficiency, and payload sway.

Findings indicate that there is a strong negative correlation between the task difficulty and operator performance. The easy condition had the smallest time to complete (64.0 s) and the largest path efficiency (0.224) whereas the hard condition had the longest time to complete (185.5 s) and the highest payload instability (40.46° maximum sway). These results confirm that difficulty-based progressive training is a useful model of developing crane operators. The project is a computing-based project which combines 3D physics simulation, real-time human-machine interface, and automated data-capture to measure and optimize operator performance — directly applying software engineering, human-computer interaction, and data analysis to a real-world safety issue.

1. Introduction

Problem, Motivation, and Significance

The work of tower crane operator is a safety-sensitive occupation that involves strict motor control, sharp sense of space, and solid judgment of the load. Traditional training systems, such as on-the-job and physical simulators, are commonly too costly to afford, not easily scalable, and unsafe to the novice operator [11, 12, 17]. Although physical environments expose trainees to occupational hazards, high-fidelity virtual simulators provide an alternative (safe, repeatable, and controlled) way of learning occupational safety and skills [15, 18]. In particular, the shift to Immersive Virtual Reality (VR) and Desktop (DT) simulators offers an affordable early-stage training solution [16]. Nevertheless, the relative efficacy of these modalities, especially with regard to depth perception and spatial disorientation, is a project question that needs controlled testing [10, 14].

Proposed Solution and Strategic Scope Shift

This senior project provides a high-quality virtual tower crane training simulator developed in Unity 2022 LTS [5], deployed in VR and Desktop platforms. First, the project was to create a real-time digital twin connected with a physical crane model, which is consistent with the new applications of Industry 4.0 [4, 8]. Nevertheless, the project experienced a strategic scope change halfway through the academic year. The physical hardware integration was postponed and relegated to future effort to give priority to experimental validity and to maximize visual and physical fidelity. This pivot was effective in removing hardware-based bottlenecks and organizational friction commonly linked to the industrial use of VR [1], which enabled the team to create a highly controlled virtual environment that is optimized to evaluate human performance in a standardized way across various difficulty levels.

Project Objectives and Contributions

The main goal of the project was to create a coherent cross-platform framework which enables the execution of similar tasks and still retains the inherent advantages of each modality. To accomplish this, the team introduced first-person immersive control in VR and extended visual support in Desktop mode, such as first-person, third-person, and movable camera perspectives. This project adds a standardized multi-challenging task framework and an automated logging framework which discloses essential safety disparities between immersive and non-immersive perspectives. These contributions offer a solid base to examine operator efficiency, path precision, and safety-related events, including near misses.

Report Organization

The rest of this report will be structured in the following way: Section 3 will cover project constraints. Section 4 discusses the background literature on virtual training and simulation. The software architecture is described in Section 5. Section 6 outlines the project execution process that is iterative. Section 7 shows the assessment plan and the quantitative findings. Lastly, Section 8 concludes the research and proposes potential future research.

2. Identification of Constraints

Some of the main constraints that affected the development, deployment, and evaluation of the VR and desktop tower crane simulations included user safety and hardware constraints. These limitations necessitated strategic trade-offs to make the project viable and the results of the experiment valid.

Health, Safety and Ethical Considerations

The greatest limitation in the VR modality was the danger of motion sickness (simulator sickness) which is a frequent health and safety issue in immersive environments. In testing, a few of the participants felt mildly uncomfortable and nauseous because of the sensory contradiction of continuous virtual locomotion. In order to reduce this health risk and provide ethical treatment of the participants, the experimental protocol was revised so that it incorporated compulsory rest periods between the easy, medium and hard levels of difficulty. This enabled the users to physically recover, avoiding physical distress and making sure that the performance data was not distorted by physiological discomfort.

Technical Limitations and Trade-offs

The computational capabilities of the VR hardware were a limitation to achieving a fair and balanced

comparison between the desktop and VR modalities. As the Meta Quest 2 is an Android standalone headset, it has much lower GPU and CPU power than a typical PC. In order to sustain a constant frame rate of more than 70 FPS — which is an absolute requirement to further prevent simulator sickness — major technical trade-offs were necessary. The team needed to optimize the 3D models, lower the total texture resolution and use aggressive texture compression algorithms (like ASTC) to the Android build. This led to a minor decrease in visual fidelity but ensured the smooth, deterministic behavior needed to log data accurately and allow the user to be comfortable.

Economic and Resource Constraints

The project enjoyed the support of the institution since the physical hardware (Meta Quest 2 headsets and testing computers) was offered by the faculty advisor. Although this removed both direct economic constraints and budget risks to the team, it in effect limited the deployment ecosystem to the capabilities of the supplied Oculus hardware, and did not allow exploration of more high-end, PC-tethered VR systems (e.g., Valve Index) which could have provided higher graphical fidelity without trade-offs.

Environmental and Sustainability

The environmental impact of the project is very positive from a sustainability perspective. The conventional training of crane operators involves the use of huge physical equipment, which uses a lot of electricity or diesel fuel, and exposes the equipment to mechanical damage and wear. Using a digital twin environment fully, the system will offer a training option that will be environmentally friendly and will significantly lower the carbon footprint and the material expenses of industrial training.

3. Background and Related Work

3.1 The Digital Twin as an Industry 4.0 Catalyst

The increasing complexity of industrial system design and simulation as well as interaction has led to the development of digital twins (DTs) as potent 3D model and simulation tools [1]. A digital twin is a virtual image of an object or a system in real life that allows simulating it, monitoring, and optimizing its performance through all the lifecycle stages [2,3]. This concept was first proposed in 2003 by Michael Grieves as part of the Product Lifecycle Management (PLM), officially adopted by NASA in the Integrated Technology Roadmap in 2010, and put into operational use on spacecraft by 2011 [3].

Digital twins have also become crucial to Industry 4.0, which combines the Internet of Things (IoT), cyber-physical systems (CPS), and artificial intelligence (AI) in order to transform industrial processes [4]. As remote-viewing platforms, predictive analytics, and fast optimization of 3D spaces, DTs are part of a fast-growing market, which is set to reach 26.46 billion by 2025 compared to 7.47 billion in 2021 [3].

3.2 Core 3D Modeling Platforms for Digital Twin Systems

The integrity and usefulness of the digital twin is largely determined by the 3D modeling platform selected to host real-time simulation and data fusion [2]. The modern digital twin workflows are dominated by three platforms. Unity is popular in visualization of simulations and industry because of its real-time data ingestion and physics-based interactions that can be achieved through C# scripting [2, 5]. Unreal Engine is particularly strong in photorealistic rendering, as well as having strong VR/AR support via C++ and Blueprint, which is a visual scripting system [2]. Blender is more of a 3D environment used to create assets which are then imported to downstream simulation engines [2].

Among them, Unity 2022.3 LTS and Unreal Engine 5 are the state of the art in high-fidelity development of crane simulators, where a rapid slew operation by an operator is accurately simulated by the physical effects of momentum and inertia of the actual machinery [4, 10].

3.3 Digital Twins for Heavy Machinery and Crane Systems

Digital twin technology is increasingly being used with sophistication on heavy industrial equipment like cranes, which work under hazardous conditions, and training in real-world conditions is expensive and unsafe [3, 6]. A typical representative harbor crane DT architecture is composed of four layers: an Equipment Layer with the physical machinery, a Data Layer with SQL-based storage, a Software Layer with Unity-based visualization of the data through C# scripts, and a Display Layer with graphical interfaces to the operator [3].

The crane systems can be virtually replicated 1:1 with high-fidelity geometric modeling in CAD tools like SolidWorks [3]. Further simulation models use meta-action theory to break down crane motions into analyzable units, enabling dynamic modeling and system identification. Hardware-in-the-loop (HIL) testing is another application of these models to test the safety and reliability of crane control software before it is

deployed [6]. The most recent studies also combine reinforcement learning in the planning of trajectories and the optimization of control in tower crane tasks [7].

One of the crucial technical considerations in any crane simulator is the careful simulation of dynamic payload oscillation — the pendulum effect that happens when the load swings during a boom swing or motion. The oscillation time depends on the hoist rope length and the acceleration of gravity, but practical simulators also need to consider the weight of the load, the wind, and the deflection of the boom [10]. These simulators are modified to be diagnostic through automated performance logging, where the performance measures include execution time, maximum load swing angle, collision, near-misses, and two-blocking [12, 13].

3.4 Immersive VR vs. Desktop-Based Simulation for Operator Training

There exists a considerable amount of literature that compares Immersive Virtual Reality Interactive Environments (IVRIE) to traditional desktop-based (DT) simulator systems when training in heavy machinery. The basic difference is the form of the presence of space and the sensory information that the trainee can receive. Whereas desktop systems offer a semi-immersive experience by displaying the information on a flat screen with a controller to navigate the system, IVRIE uses Head-Mounted Displays (HMDs) to offer the user a human scale of 1:1 and a 360-degree viewshed [11].

It has always been shown that the accuracy of spatial perception depends directly on the degree of immersion. It has been found that participants on desktop systems also take virtual spaces systematically larger because they lack egocentric depth cues — the possibility to use their own bodies as a reference point in space [11]. Historically, VR systems had an average ratio of underestimation distance of about 73.48 percent, but the increase in pixel density, field of view, and weight of the HMD within the years 2018–2026 has significantly reduced this [15]. Presence — the psychological sense of “being there” — is a key idea in training effectiveness because high presence promotes the situational awareness and spatial cognition involved in the real crane operation [11, 16].

Operator precision is also greatly dependent on visual perspective. The simulators provide the first-person in-cabin perspective, third-person outside perspective, as well as mobile camera perspective. Studies show that the availability of high-view capabilities during the initial training can extend reaction time

up to 43 percent in reaction to unforeseen events, but such aids have to be gradually withdrawn as the trainee advances to the actual equipment to facilitate transfer of training [10, 12]. Teleoperated crane performance is further improved with the addition of haptic force-feedback and auditory modalities, but research reports a trade-off: multimodal aids can make the task of the operator faster, but sometimes may increase the frequency of moderate collisions because the operator becomes over-reliant on the assistive cue, and cannot develop independent spatial judgment [14].

3.5 Integration of Immersive and Intelligent Technologies

Digital twins, VR, haptics, and AI merge to transform human-machine interaction in new applications in Industry 4.0 [4, 8]. VR applications developed in Unity, with libraries like the XR Interaction Toolkit and Oculus Integration SDK, can be used to visually simulate crane telemetry in real-time and control physical equipment remotely [4]. Haptic devices provide haptic feedback, making remote machine interaction more realistic and facilitating accurate teleoperation [9]. AI generative models trained on CAD data are able to simplify the creation of digital assets, understand natural language operator commands, and allow prediction of what-if simulation [4].

A recent experiment examining lightweight mobile augmented reality (AR) in crane safety training on two-dimensional load-chart perception and spatial awareness indicated that dual-mode 3D visualization allows novice operators to gain the benefits of immersive technology not just in fully equipped VR stations, but also in mobile devices available on construction sites [13]. Also, it has been demonstrated that Digital Twin-Driven (DTD) predictive models can be used to predict degraded lifting capacity of old cranes with high precision ($MSE = 0.2253$), which allows operators to train on a virtual duplicate of the machine they are going to operate in its current structural condition [7].

3.6 Industrial Training Standards and Learning Transfer

The eventual test of the worth of a virtual training system is the transfer of learning — the extent to which the skills learned in simulation are applied accurately to real world operation. The quantitative data is overwhelming: companies which have introduced high-fidelity VR training report a decrease in time-to-competency of new employees by 25–40% and up to a 90% loss of training related to dam-

aged equipment and infrastructure [16, 17]. The steel and port sectors are especially educative in this regard: ArcelorMittal Gent managed to decrease pick-ups during physical handling by 30 percent, moving the training to a virtual setting, and international port authorities have recorded a 50 percent decrease of real-life accidents since simulators became part of the onboarding [17].

The design of effective training curricula is based on Multi-difficulty tasks that are evaluated based on either Hierarchical Task Analysis (HTA) or Cognitive Task Analysis (CTA), and advance trainees to load stabilization and emergency response situations [18]. The Occupational Safety and Health Administration (OSHA) standard 29 CFR 1926 Subpart CC specifically acknowledges virtual platforms as a technique to develop hand-eye coordination to construction machinery and certification programs such as the Specialist in Safety and Health (SSH) track necessitate course work directly pertinent to crane operation and machine protection [17, 18].

3.7 Challenges in Implementing VR-Based Digital Twins

Nonetheless, there is still much to be done to bring VR-powered digital twins to traditionally conservative industries, including technological and organizational barriers [1]. Hardware incompatibility, the fast obsolescence of VR devices, limited bandwidth in data transfer, and VR simulator sickness due to latency or unstable frame rates are examples of technological barriers [1]. On the organizational level, the problems of internal coordination, the mismatch of priorities between different departments, and the opposition of operators to new technologies drag the adoption. Maintenance providers tend to wait until there is a sufficiently large number of users before investing in the skills required, hence adding to a compounding delay across the field [1].

Economic factors also play a great role. The initial capital investment in a high quality crane simulator is offset by downstream savings in equipment wear, instructor expenses, site idle time, and risk of accidents. Steel Sim VR projects that 75 operators can be trained in a single simulator in a year with 25% savings in time spent on the production floor; tens of thousands of euros are saved in lost material and production delays, respectively, in large scale operations like those at Tata Steel or Algoma Steel [17].

4. Project Approach

4.1 Solution Description

The project provides a virtual tower crane training simulator of high fidelity deployed in two similar modalities: Virtual Reality (VR) and Desktop using Unity 2022 LTS. The essence is to offer a low risk, affordable and sustainable digital platform to train crane operators and test human performance in its initial phases. The two modalities are similar in terms of the crane model underlying them, the task structure, and the difficulty framework, which allows a controlled comparison of operator behavior in similar situations.

The simulator recreates the key functional elements of tower crane control: trolley travel along the track, jib rotation (boom slewing) and hoisting and lowering with the cable length control, resulting hook motion, and dynamic payload oscillation. The combination of these four degrees of freedom (DOF) is the kinematic envelope of the crane, on which all the work activities are performed and all their performance is recorded.

4.2 Software Architecture

The architecture is built on top of a common simulation core and interface-layers that are modality-specific:

Shared Core: The crane model, physics engine, task scenario logic and the automated logging framework are found in a common Unity scene. The geometry of all tasks such as pick-up area, transportation routes, and drop targets is also such that both modalities have equal layout opportunities across all levels of difficulty.

VR Mode: The VR mode is compatible with the Meta Quest 2 headset and is physically directed with Oculus Touch controllers. The player is placed in a first-person immersive view of the virtual crane cab where he is able to engage with the world around him as an operator. Both the XR Interaction Toolkit and Oculus Integration SDK were merged to facilitate VR device communication and mapping controller input. Tethered development and testing was done in the Meta Quest Link pipeline.

Desktop Mode: The Desktop mode is controlled with keyboard inputs and has visual support outside of the VR modality. Besides a first-person cabin view, the Desktop mode offers a third-person view, and a free-moving external camera. It also has a virtual dashboard-based cabin view to aid in hook and cargo monitoring. These extra perspectives enable users to inspect the hook placement, estimate relative dis-

tance, and refine cargo placement more specifically when picking up and delivering the cargo to the destination, a significant design benefit of spatial accuracy.

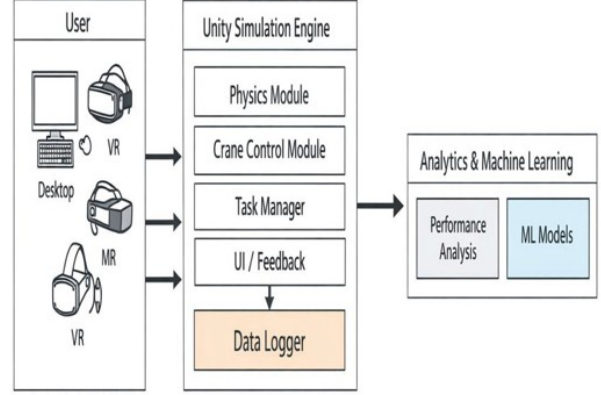


Figure 1: System Architecture

4.3 Task Scenarios

The simulator has pick-and-drop scenarios that are arranged into three difficulty levels: easy, medium and hard. All levels involve having the operator complete a series of cargo pickups, controlled horizontal transportation and proper placement at a specified drop point. Difficulty is gradually added by making the space more limited, adding obstacles and decreasing the margins of the workspace around the target area. Each of the three difficulty levels has the same basic task objective, thus making performance measures directly comparable across conditions.

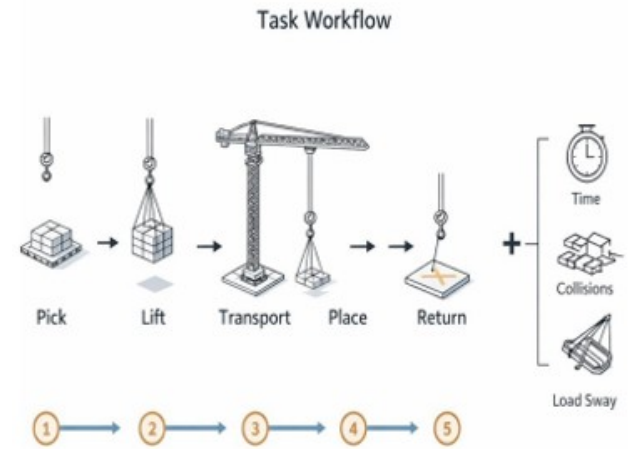


Figure 2: Task Workflow

4.4 Automated Logging Framework

A major system architecture element is its inbuilt automated logging pipeline. Two complementary types

of information are documented in the framework during each trial:

The data sent by the state is continuous and sampled at consistent time intervals and consists of: simulation time, trolley position, jib angle of rotation, cable length, payload oscillation angles around two axes and the three-dimensional position of hook. The variables can be completely reconstructed to obtain crane motion and load behavior over time.

Event-based data (event data, event markings, or moment marks) identify separate operational points in time, and in some cases, place as well. Recorded events are: Game Started, Cargo Picked Up, Near Miss, Cargo Delivered and Game Ended.

Exported data is in CSV and JSON format to analyze after the trial. Based on these raw logs, the following trial-level performance measures are obtained: 3D hook path length, straight-line start-to-end displacement, straight-line task completion time, path efficiency (ratio of straight-line displacement to actual path length), maximum and mean payload angle of sway, hook speed profile, pickup success, delivery accuracy, and frequency of near-miss.

4.5 Data Processing Pipeline

Raw log files are read from header to the end of each trial. Time is normalized, such that every trial starts at time $t = 0$ s and angular variables are converted to degrees. The sway along the two swing axes is pooled together to form a single effective sway to act as a single measure of load stability. The changes in positional state between successive time steps of the trolley, cable and hook are used to derive velocity and acceleration which aid in kinematic analysis of the quality of operator control.

4.6 Third-Party Components

The following external components were integrated into the system without internal implementation:

Table 1: Third-Party Components

Component	Purpose
XR Interaction Toolkit	VR input handling and controller mapping
Oculus Integration SDK	Meta Quest 2 device support
Meta Quest Link	Tethered VR development and testing
Unity 2022 LTS	Core simulation engine, physics, scripting (C#)
Tower Crane Assets	Ready assets for integration

4.7 Tools and Project Management

GitHub was used to manage version control, which allowed team members to work on them simultaneously. The scheduling of tasks based on sprint was managed through Jira which gave insight into the progress, blockers, and milestone tracking during the two semesters. The Universal Render Pipeline (URP) of Unity was chosen as the rendering backend after some premature pipeline wrangles were experienced during development.

4.8 Team Structure

The team split tasks into three main workstreams: simulation and physics development (crane kinematics, task geometry, difficulty design), interface and modality implementation (VR controller integration, Desktop camera system), and data infrastructure (logging framework, CSV/JSON export pipeline, metric extraction scripts). Synchronization meetings were conducted regularly to align the cross-stream dependencies especially of shared scene architecture and definition of logging events. Physics instability was solved by collaborative debugging to render pipeline problems that arose during development.

5. Project Execution

During the previous two semesters, the project shifted its focus towards initial research and prototyping to the full implementation of the system and user testing. This stage demanded both adaptive design and technical problem-solving as well as effective teamwork.

Teamwork and Division of Responsibilities

Each of the team members had their responsibilities split according to their technical strengths to guarantee effective progress:

- **Yestay** assumed the leadership position where he coordinated team works, arranged meetings, and monitored deadlines. He was also a co-leader in the design of the experiment and user testing.
- **Shynggys** designed the mechanical interactions, with the physics and kinematics (four degrees of freedom) of the crane to make the movement and load sway realistic, and worked on the data logging system.
- **Almas** handled the visual and immersive elements, taking the responsibility of the 3D modeling of the crane and the environment, the VR integration, and the installation of the XR Interaction Toolkit in the Meta Quest 2.

- **Bekzhan** worked on user interfaces and data collection, creating the UI/UX, the main menu, and the logic of difficulty selection, and the *CraneLogger* system to log the telemetry to .csv files.
- **Saken** was in charge of input settings of the desktop version and establishing keyboard controls and camera systems and was actively involved in the planning and execution of the user experiments.

Design Choices and Project Development

First, the project scope was thought of as a rigid separation of point of view: VR would be first person view (FPV) and the desktop would be third person view (TPV). But as the team went through the iterative design process, they realized that such an asymmetry will confound the human performance evaluation. The last design was extended to FPV and TPV on both platforms. The system was scaled to 12 different operational scenarios in combination with the three difficulty tiers. This normalized the visual stimuli in all modalities so that the variation in performance could be directly attributed to the input modalities and not the camera angles.

Technical Problem-Solving

During the development process, the team has faced and solved a number of intricate engineering problems:

Build Pipeline and Gradle Deadlocks: During the deployment to the Meta Quest 2, the Android build pipeline would often fail because of Gradle cache locks (registry.bin lock errors). Background Java applications were not killing properly and stopped compiling .apks. The solution that the team found to this was to use custom command-line overrides to kill orphaned processes and to purge the cache of the builder systematically.

Input Architecture Conflicts: The problem of integrating VR and desktop modalities in the same Unity project resulted in input handling conflicts. The XR Interaction Toolkit needed the “New Input System Package” and the old desktop scripts used the older Input Manager. The team reduced this by separating the input logic and dynamically changing the backend settings depending on the target platform.

Scoped Storage and Telemetry Extraction: The *CraneLogger* system is a system with detailed .csv exports. Whereas regular paths were functional on Windows, Android security policies of the Meta Quest 2 prevented file writing to system folders without warning. In order to retain the data, the team adopted platform-specific directory routing, and used

Android Debug Bridge (ADB) and SideQuest to directly extract logs through the headset.

Testing Implementation

The team was able to switch to data collection at the culmination of the second semester. The pilot was carried out involving a group of 5 participants. All participants were taken through the entire range of 12 scenarios. The team performed hardware configurations, oriented users on safety measures, launched telemetry recording, and received qualitative feedback via standardized SUS and NASA-TLX questionnaires during these sessions. Every quantitative and qualitative information was extracted successfully to be evaluated.

6. Evaluation

6.1 Experimental Setup

Five subjects participated in a study. All the participants went through all twelve conditions, resulting in 60 recorded sessions. Order of conditions was balanced out among the participants to alleviate order effects (e.g., Subject 1 started with Desktop FPV; Subject 2 started with VR TPV; Subject 5 started with VR FPV). But all subjects had to commit to one modality and one mode at a time and go through level 1 (easy), 2 (medium), and 3 (hard) in that order.

A short introduction to the control scheme and the crane operator’s aim was given to participants before formal trials. None of them had any previous experience in the crane-operating profession. Qualitative observations were recorded by the examiner during and after the session.

6.2 Analysis Approach

Exactly five sessions ($N = 5$, one per participant) each represented each of the twelve platform-viewpoint-level conditions. Each performance variable was calculated using descriptive statistics (mean, standard deviation) according to each condition. Since the sample size ($N = 5$) is very small, no inferential statistics are reported; the measures are reported in the form of planned evaluation metrics, whose trends can be easily explained by the research questions and can be quantified by further studies with larger sample sizes in the future.

6.3 Quantitative Results

6.3.1 Task Completion Time

The duration of completing tasks grew monotonically with the level of difficulty in all of the four conditions, and the three-level task structure did generate a true gradient of operational complexity. The completion

times scaled with task difficulty. There was more variability in FPV conditions at Level 3 than in TPV conditions, which suggests that FPV-based operation is more sensitive to individual variation.

Table 2: Mean task completion time in seconds (SD) by condition and level.

Condition	L1 Mean (SD)	L2 Mean (SD)	L3 Mean (SD)
Desktop FPV	137.3 (55.5)	325.3 (176.5)	516.1 (274.9)
Desktop TPV	155.3 (95.2)	246.5 (132.8)	322.4 (40.9)
VR FPV	100.2 (59.3)	256.4 (72.2)	447.5 (211.3)
VR TPV	171.4 (80.0)	317.3 (118.4)	385.4 (92.3)

6.3.2 Near-Miss Count

The most diagnostically informative variable in this dataset is the near-miss count. At Level 1 and Level 2, there were no significant differences in the number of near misses, except for VR TPV Level 2, where all sessions demonstrated a higher number of near misses compared to other conditions. However, at Level 3, VR FPV stopped being similar to all other scenarios: the average number of near misses was 86.0 (SD = 27.3), and the number of events ranged between 51 and 126 in separate sessions. By contrast, the number of near misses for VR TPV, Desktop TPV, and Desktop FPV remained below 25. The combination of immersive VR and high task complexity in a first-person view directs the blame to this finding. However, Desktop FPV also stood out from the rest, with twice as many mean near-misses and high SD. This is because in FPV, the operator’s spatial reference frame is closely connected with the hook camera, which can be rotated 180 degrees, leading to a complete loss of spatial orientation, as claimed by all five subjects, which appears to lead to a higher number of near misses.

Table 3: Mean near miss count (SD) by condition and level. VR FPV Level 3 is a clear outlier.

Condition	L1 Mean (SD)	L2 Mean (SD)	L3 Mean (SD)
Desktop FPV	2.2 (0.8)	6.0 (1.0)	22.8 (13.9)
Desktop TPV	1.8 (0.8)	6.0 (1.6)	11.2 (4.6)
VR FPV	2.0 (1.2)	5.2 (0.4)	86.0 (27.3)
VR TPV	2.4 (1.1)	9.2 (1.5)	11.2 (2.2)

6.3.3 Collision Count

At Level 1 and Level 2, there were no or minimal collision events under all conditions, indicating the reduced spatial demands of the corresponding tasks. In Level 3, it became a distinguishing variable. The mean collision count was highest at Desktop FPV, but this was greatly affected by one outlier session (24 collisions). Without this outlier session, the remaining four sessions have a mean of 3.75 collisions. The FPV conditions resulted in a higher number of

collisions when compared to their TPV counterparts at hard difficulty in both platforms, which is in line with the near-miss pattern and the orientation loss reported qualitatively. The high standard deviation of the Desktop FPV Level 3 (SD = 9.70) should be treated with caution when making interpretations of the mean and is indicative of the high impact of individual strategy and spatial awareness at that level.

Table 4: Mean collision count (SD) and Level 3 range by condition.

Cond.	L1	L2	L3	L3 Range
Desktop FPV	0.0 (0.00)	0.6 (0.89)	7.0 (9.70)	0–24
Desktop TPV	0.2 (0.45)	0.2 (0.45)	1.4 (1.14)	0–3
VR FPV	0.0 (0.00)	0.4 (0.55)	3.0 (2.45)	0–6
VR TPV	0.6 (0.89)	1.0 (0.71)	2.2 (1.64)	1–4

6.3.4 Hook Path Length

All conditions showed an increase in hook path length with difficulty. In Level 1, paths of 180–260 m were formed under all conditions. At Level 3, Desktop FPV produced the longest mean path, followed by VR FPV, Desktop TPV, and VR TPV. The high variability of FPV at Level 3 is a reflection of its task time, near-misses, and collision variabilities. TPV generated the most consistent courses in all the levels, indicating the most consistent movement plans.

Table 5: Mean hook path length in metres (SD) by condition and level.

Condition	L1 Mean (SD)	L2 Mean (SD)	L3 Mean (SD)
Desktop FPV	257.2 (50.8)	725.5 (336.7)	970.2 (487.5)
Desktop TPV	200.4 (34.9)	481.3 (82.4)	579.7 (47.7)
VR FPV	180.4 (63.7)	500.8 (53.5)	680.3 (203.2)
VR TPV	211.1 (37.6)	503.3 (42.8)	529.2 (41.6)

6.3.5 Cargo Oscillation

Mean oscillation magnitude did not vary much according to condition, but it was in the range of about 0.069–0.102 rad across all twelve cells. There was no significant difference between VR and Desktop or between FPV and TPV. This implies that the physical oscillation dynamics of the crane were perceived similarly across modalities and perspectives, and that variations in the number of near misses and the number of collisions cannot be explained by the fact that, in one of the conditions, the participants were swinging the load more violently. Because the range of values is narrow, the oscillation magnitude is not tabulated separately; the descriptive summary above is sufficient to establish that crane physics behaved consistently across all conditions.

6.3.6 Task Success Rate

The success of the task (Cargo Delivered logged) was 100% in all 60 sessions. There was one session (VR FPV Level 1) in which no “Cargo Picked Up” event was recorded, but a “Cargo Delivered” event was recorded, which is explained by an inconvenient logging edge case at the start of the session as opposed to a failure of the task. None of the participants were unable to deliver the cargo in any of the conditions or difficulty level, which proved that all four simulator configurations are operationally complete at the present task design.

It is important to note that the assessment had a risk of health issues, such as dizziness, nausea, soreness of the eyes, and headache, especially under VR conditions. These symptoms were also reported throughout the participant group to differing extents and in other cases, so severe that they posed a risk of disrupting the assessment. In all cases, complications were prevented by proactively keeping track of the well-being of participants during the sessions and ensuring that they had a rest break before continuing. This protocol was critical to the completion of 100% of the tasks and should be viewed during any future replication or generalization of this study.

6.4 Qualitative Results

6.4.1 Spatial Disorientation from Hook Camera Rotation

The five participants all complained that there was disorientation due to the rotation of the hook camera in the FPV mode. In the case of the hook camera being turned 180 degrees, the spatial frame of reference of the operator was turned upside down, which would result in a disconnect with their inputs. For example, if an operator wants to rotate the crane boom clockwise and presses the corresponding button, they then see that the crane rotates counterclockwise from their perspective. This further fueled health complications in VR. Subject 1 explained this specifically and was forced to have a lengthy rest period because of VR-related fatigue after the high-difficulty VR sessions. Subject 5 got nauseated after finishing the last difficulty level. Subject 4, in particular, did not like unintentional camera rotation due to collisions in FPV mode, which directly relates the camera-flip disorientation to the high number of collisions recorded in FPV conditions in Level 3. The fact that this complaint is common to all subjects makes rotation of hook cameras the most reliable usability problem that has been found in the study, and its remedy the most urgently actionable design solution.

6.4.2 Difficulty Locating Cargo and Destination Points

The five participants said that they found it challenging to find cargo and destination points in the tasks. Importantly, this issue was limited to FPV conditions — Desktop FPV and VR FPV — and was not described in TPV conditions. The small field of view and the lack of a map or spatial guidance overlay in FPV meant that targets were required to be searched by moving the crane, frequently with no success on the first try. This is clearly manifested in the increased and more fluctuating hook paths in FPV conditions at Level 3, where the inefficiency of navigation is added to the difficulty of control. The broader, raised view of TPV offered enough space context so that the participants would always be able to find the targets without further instructions.

6.4.3 Viewpoint Preference and Prior Experience

Subject 2 did not like Desktop mode because he considered it too artificial and liked VR with its immersive quality. Subject 4, in its turn, was more inclined to TPV by the fact that it did not experience the accidental camera rotations in FPV mode, and did not hesitate in the Desktop TPV condition, which is in line with the low variability in the task time in this condition (Table 2, SD = 40.9 s at Level 3). Both Subject 3 and Subject 2 had extensive gaming experience, and both adapted more quickly to the control scheme of the simulator, implying that prior exposure to virtual environments hastens the acquisition of skills in both modalities.

6.4.4 Fatigue and Physical Comfort

VR conditions were characterized by a significant amount of physical and cognitive fatigue in comparison with Desktop conditions. Subject 1 experienced more fatigue in VR compared to Desktop and needed a long break after VR sessions. Subject 5 felt nauseated following the last level of VR. Subject 3 complained of eye soreness. None of the participants complained of discomfort in the Desktop mode, other than the initial control scheme adjustment period (Subject 4 early trouble with Desktop FPV controls). These findings support the note on task success rate concerning health monitoring: VR simulator studies need to be actively controlled in terms of welfare, and subjective measures of comfort and workload should be included in future research.

6.5 Addressing Research Questions and Objectives

6.5.1 Problem Statement and Aim

The thesis planned to execute a high-fidelity tower crane simulator both in VR and Desktop modes,

and developed a multi-difficulty pick-and-drop task evaluation-ready framework. This objective was confirmed by the assessment: the 60 sessions in 5 participants were completed successfully with both modalities and all three difficulty levels, and the data were automatically recorded in each of the sessions. The logging system generated half a dozen derived performance measures on which this analysis is based, without any form of post-processing of the simulator itself.

6.5.2 Research Objective 1 — Consistent Task Logic Across Platforms

The necessity to provide similarity in the logic of the tasks of VR and Desktop is supported by the fact that in each of the modalities, the success rate of the tasks is 100 percent. The entire pick-and-drop sequence could be undertaken in every session. The difficulty was kept constant as well: the time to complete, the number of near misses, and the number of collisions all rose with level in the four conditions (Tables 2, 3, and 4). This confirms that the task structure underlying was identical in both platforms.

6.5.3 Research Objective 2 — Modality-Specific Viewing Conditions

The findings indicate a definite interaction: the point of view is of great significance in VR compared to Desktop. The difference between FPV and TPV in near misses at Level 3 was 11.6 events in Desktop (22.8 vs. 11.2), and in VR, the difference was 74.8 events (86.0 vs. 11.2). Collisions in Level 3 showed 5.6 more collisions with Desktop FPV than with Desktop TPV, and VR FPV than with VR TPV — the near miss variable is more sensitive here. Desktop FPV, softened by the impersonality of using the keyboard as a control device and the lack of health complications, maintains a level of comfort that VR FPV lacks. The most consistent and efficient paths were observed in Desktop TPV and VR TPV between levels, and more variation and longer paths were exhibited in FPV conditions with an increase in difficulty.

6.5.4 Research Objective 3 — Logged Variables for Analysis

According to the statistics of 60 sessions, the following conclusions can be made on the appropriateness of logged variables:

- The near-miss count is the most discriminative safety variable and clearly distinguishes the problematic VR FPV condition at Level 3, compared to any other.
- Collision count offers supplementary safety data

that supports the near miss pattern, but due to its large variance in Desktop FPV Level 3, it is not reliable on its own with $N = 5$.

- Task completion time is the most important efficiency measure because it is efficient and scales well with difficulty.
- Hook path length is a control accuracy and redundancy measure that complements task time by separating fast-but-indirect and slow-but-direct strategies.
- The average magnitude of oscillation was similar in all the conditions and showed that the crane physics is represented equally. It can distinguish strategies using a bigger sample or trained operators.

6.5.5 Validation of the Project Solution

The assessment confirms the simulator as a practical study and training tool in three aspects. To begin with, the system was able to operationalize all design goals: it executed the same task logic in both modalities, all perspectives were functional, all three levels of difficulty generated different performance profiles, and the logging architecture was able to record all the variables of interest without loss in 60 sessions. Second, the generated data displayed interpretable and theoretically sensible patterns: the near misses, collisions, time to finish the path, and the length of the path all grew predictively with difficulty, and both the VR FPV near miss and collision patterns can be directly explained by the disorientation mechanism that was observed in all five participants. Third, the system demonstrated a substantively significant design result, namely that FPV at high task complexity results in safety-relevant orientation loss not present in TPV. This is exactly what the simulator was intended to allow: a cross-platform, cross-viewpoint comparison.

These findings combined prove that the simulator is an evaluation-ready framework and fulfills all three research objectives. The small sample size ($N = 5$), the use of no formal inferential statistics, and the absence of subjective measures like workload or presence ratings are seen as limitations presented as future work directions.

7. Conclusion and Possible Future Work

Conclusion

This senior project was able to design, implement, and assess a high-fidelity tower crane training sim-

ulator in VR and Desktop modalities. The team developed a strong quantitative operator evaluation framework by focusing on simulation fidelity and performance measurement in controlled conditions, as opposed to real-time hardware telemetry. The automated logging pipeline was very efficient as it logged continuous variables and discrete events in 60 trials with 100% success rate in the task.

The experimental findings show that there is an apparent negative correlation between operational efficiency and task difficulty. Most importantly, the assessment showed that there are serious safety implications in terms of viewpoint choice. The immersive VR First-Person View (FPV) at the highest difficulty level resulted in a dramatic increase in near misses, with a mean of 86.0 events, which is about 7.7 times the number of near misses in any other condition. The participants said that when they attached their spatial reference frame rigidly to the rotating hook camera they lost all spatial orientation. On the other hand, Desktop mode Third-Person View (TPV) led to much fewer collisions and more predictable path lengths. This is consistent with the recent literature which indicates that although VR offers high presence, visual constraints in depth perception and field of view may hinder the execution of complex tasks [10, 14]. These results affirm that although immersive VR is most effective in presence, additional aids to viewing are essential in ensuring spatial awareness when performing complicated maneuvers.

Possible Future Work

The existing simulator serves as a fully functional platform and provides a basis for several advanced extensions:

- **Physical Crane Synchronization:** Following the strategic scope shift, one of the primary avenues for further integration into a full-fledged cyber-physical system is to connect the digital simulator with a physical crane model utilizing real-world telemetry [3, 6].
- **Smart Path Planning:** The application of Reinforcement Learning (RL) agents may automate the process of planning optimal crane trajectories, providing AI-based training suggestions and predictive collision avoidance [7].
- **Haptic and Sensory Integration:** Force-feedback devices could be added to simulate the physical sensations of payload sway and cable tension, bridging the tactile gap in existing VR training [9].
- **Dynamic Environmental Challenges:** Future versions should introduce dynamic weather conditions, such as wind shear, rain, and low visibility, to test operator resilience in high-stakes, real-world scenarios.
- **Expanded Human-Subject Testing:** Increasing the participant pool will enable robust inferential statistical analysis, validating the current performance metrics across a broader demographic of operators.

References

- [1] M. Saghafian, K. Laumann, and M. R. Skogstad, "Organizational challenges of development and implementation of virtual reality solution for industrial operation," *Frontiers in Psychology*, 12:704723, 2021.
- [2] R. Green, "3D Modeling Digital Twins: Unity vs Unreal vs Blender," NeticSpace, n.d.
- [3] "Design of Digital Twin System for Harbor Cranes Based on Unity 3D," AWS, n.d.
- [4] A. Demirbas and W. Xu, "Real-time virtual reality visualization and control for digital twin applications in industry 4.0," *Journal of Student Research*, 13(3):1–10, 2024.
- [5] Unity Technologies, "VR Development Pathway – Learn Unity VR Game Development," Unity Learn, 2025.
- [6] Controllab, "Steel Crane Digital Twin," Controllab, 2025.
- [7] TUM School of Computation, Information and Technology, "Design and Simulation of Tower Crane 3D Lift Task Integrated with Reinforcement Learning," Department of Computer Engineering, n.d.
- [8] C. Yang, "Framework for Virtual Reality Digital Services Leveraging Digital Twin-Based Crane," Master's thesis, Aaltodoc, 2021.
- [9] Wikipedia Contributors, "Haptic Technology," Wikipedia, n.d.
- [10] "Understanding the Effects of Virtual Reality System Usage on Spatial Cognition and Presence," *Sustainability*, 14(16):10326, MDPI, 2022.
- [11] R. Nilsson and A. Sayfoor, "VR-Based Crane Simulator Training: Evaluation of Learning Outcomes and Skill Transfer," Master's thesis, DiVA Portal, 2023.

- [12] CM Labs, “Your Crane Simulator Buyer’s Guide,” CM Labs Simulations Inc., n.d. Available: <https://cm-labs.com/en/resource/your-crane-simulator-buyers-guide/>
- [13] “Lightweight Mobile AR for Crane Safety Education: Dual-Mode 3D Visualization to Enhance Load-Chart Interpretation and Spatial Awareness,” ResearchGate, n.d.
- [14] “Assessing Depth Perception in VR and Video See-Through AR: A Comparison on Distance Judgment, Performance, and Preference,” *IEEE Transactions on Visualization and Computer Graphics*, 2024.
- [15] “Analysis of the Effectiveness and User Experience of Employing Virtual Reality to Enhance the Efficacy of Occupational Safety and Health Learning for Electrical Workers and Graduate Students,” *PLoS ONE / PMC*, 2024.
- [16] “Effectiveness of VR Crane Training with Head-Mounted Display: Double Mediation of Presence and Perceived Usefulness,” *Automation in Construction*, ResearchGate, 2020.
- [17] “Steel Sim VR Crane Simulator Training FAQ,” SteelSimVR, 2025. Available: <https://steelsimvr.com/2025/10/07/steel-sim-vr-crane-simulator-faq/>
- [18] “A Review on Virtual Reality Skill Training Applications,” *Frontiers in Virtual Reality*, 2:645153, 2021.