







Article

Preliminary Design Guidelines for Evaluating Immersive Industrial Safety Training

André Cordeiro ^{1,*}, Regina Leite ², Lucas Almeida ³, Cintia Neves ⁴, Tiago Silva ⁵, Alexandre Siqueira ⁶,
Marcio Catapan ³ and Ingrid Winkler ^{4,7,*}

- ¹ Institute of Pharmaceutical Technology, Oswaldo Cruz Foundation—Fiocruz, Rio de Janeiro 21040-360, Brazil
² Construction Department, Federal Institute of Bahia, Salvador 40110-150, Brazil; regina.leite@ifba.edu.br
³ PGGDDesign, Federal University of Paraná, Curitiba 80320-230, Brazil; lucas.gregory@ufpr.br (L.A.); marciocatapan@ufpr.br (M.C.)
⁴ Institute for Science, Innovation and Technology in Industry 4.0 (INCITE INDUSTRIA 4.0), SENAI CIMATEC University, Salvador 41650-010, Brazil; cintia.neves@fieb.org.br
⁵ Department of Mechanical and Industrial Engineering (DEMI), NOVA School of Science and Technology (FCT), NOVA University Lisbon, 2829-516 Caparica, Portugal; tan.silva@fct.unl.pt
⁶ College of Engineering, University of Florida, Gainesville, FL 32611, USA; agomesdesiqueira@ufl.edu
⁷ Advanced Knowledge Center in Immersive Technology (AKCIT), Goiania 574001-970, Brazil
* Correspondence: andre.cordeiro@fiocruz.br (A.C.); ingrid.winkler@doc.senaicimatec.edu.br (I.W.)

Abstract

This study presents preliminary design guidelines to support the evaluation of industrial safety training using immersive technologies, with a focus on high-risk work environments such as working at height. Although virtual reality has been widely adopted for training, few studies have explored its use for behavior-level evaluation, corresponding to Level 3 of the Kirkpatrick Model. Addressing this gap, the study adopts the Design Science Research methodology, combining a systematic literature review with expert focus group analysis to develop a conceptual framework for training evaluation. The results identify key elements necessary for immersive training evaluations, including scenario configuration, ethical procedures, recruitment, equipment selection, experimental design, and implementation strategies. The resulting guidelines are organized into six categories: scenario configuration, ethical procedures, recruitment, equipment selection, experimental design, and implementation strategies. These guidelines represent a DSR-based conceptual artifact to inform future empirical studies and support the structured assessment of immersive safety training interventions. The study also highlights the potential of integrating behavioral and physiological indicators to support immersive evaluations of behavioral change, offering an expert-informed and structured foundation for future empirical studies in high-risk industrial contexts.

Keywords: virtual reality; industrial training; sustainable production; design science research; design guidelines



Academic Editors: Antony Bryant, Kitti Puritat, Kannikar Intawong and Wirapong Chansanam

Received: 19 June 2025

Revised: 21 August 2025

Accepted: 26 August 2025

Published: 1 September 2025

Citation: Cordeiro, A.; Leite, R.; Almeida, L.; Neves, C.; Silva, T.; Siqueira, A.; Catapan, M.; Winkler, I. Preliminary Design Guidelines for Evaluating Immersive Industrial Safety Training. *Informatics* **2025**, *12*, 88. <https://doi.org/10.3390/informatics12030088>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Sustainable Development Goal 8 (SDG 8) [1] focuses on ‘inclusive and sustainable economic growth, employment, and decent work for all’, placing occupational safety and health at the core of its agenda. Among the leading causes of fatal workplace accidents are falls from height, which remain particularly critical in construction, maintenance, and industrial sectors. In this context, improving the design and evaluation of safety training for high-risk scenarios, such as working at height, represents a concrete contribution to SDG 8.

According to global estimates, nearly 3 million people die from work-related accidents and diseases each year, with 330,000 fatalities specifically attributed to occupational accidents, and hundreds of millions of nonfatal injuries. In Brazil, falls represented 14.17% of the workplace fatalities over the last ten years [2,3].

In industries that operate in scenarios with operational risks, accidents can lead to loss of life, significant financial costs, and severe social or environmental consequences. It is therefore critical to understand how to assess the effectiveness of safety training, specifically how well it prepares trainees for real-world situations (known as training transfer) [4], when attributes such as problem-solving and analytical skills are successfully transferred and applied to the daily lives of professionals [5,6].

The Kirkpatrick Model, considered the gold standard for training evaluation, has four levels through which training can be assessed: (1) Reaction, (2) Learning, (3) Behavior, and (4) Results [5,7]. Level 3 of the model evaluates the behavior change of participants after training, measuring whether trainees apply the knowledge acquired during training in their daily work activities. This level examines the transfer of learning from the training environment to the workplace, considering whether participants have changed their practices and attitudes according to what was taught.

However, few training programs are evaluated according to the third level of the Kirkpatrick Model, regarding the actual behavior change of the trained professional. Currently, there are significant practical and methodological challenges to conducting effective evaluations of behavioral changes in the workplace, which are considerably more complex than simply measuring reaction/satisfaction (level 1) or knowledge acquisition (level 2) [5,7,8]. Behavior-focused evaluations within the workplace to determine whether trainees are actually applying what they have learned on the job require sophisticated methods such as direct observation, interviews, or detailed questionnaires applied to supervisors and colleagues, demanding specialized skills in research methodology and data analysis, competencies that are rarely available internally within organizations [4,9]. Furthermore, implementing a level 3 evaluation can be costly and time-consuming, as it requires continuous monitoring of trainees to observe how they apply their new skills at work.

Given these constraints, there is a growing interest in identifying innovative strategies capable of addressing such limitations. In this regard, immersive technologies have emerged as a promising alternative to overcome the practical and methodological challenges associated with behavior-level evaluations. Few studies, even those related to native immersive training in VR, have analyzed methods and instruments to assess training effectiveness, especially in engineering [4]. Additionally, among the studies that evaluate training outcomes, few reach level 3 of the Kirkpatrick model [4–6,8], with most authors limiting their evaluation to levels 1 or 2. More restrictively, of the few studies that reach level 3 of the Kirkpatrick model, most limit themselves to traditional assessment methods, such as interviews, questionnaires, knowledge tests, and practical tests, falling far short of what the range of sensors embedded in head-mounted displays (HMDs) allows.

Thus, considering that few training programs are evaluated regarding the actual behavior change of trainees (Level 3 of the Kirkpatrick Model) [4,8], which is considerably more complex than simply measuring satisfaction (Level 1) or knowledge acquisition (Level 2), the contradiction is that having a positive result at Level 1 does not guarantee knowledge and skill acquisition, as the data collected only reflect the reaction and overall experience, such as satisfaction, enjoyment, etc., with the given training [5,7,8]. Paradoxically, it is possible for a professional to undergo training for a high-risk occupation to evaluate a training experience as interesting and enjoyable, meaning the training is well-rated at Level 1, but not incorporate this learning as a behavior change in their daily work practices. When faced with a critical situation, they may not have the problem-solving knowledge

and could suffer accidents that may result in loss of life. Since few studies have investigated the application of immersive technologies to evaluate the transfer of learning to behavior, and the integration of multi-sensors into head-mounted displays is recent, there is a gap in knowledge about this topic.

There are also open questions regarding the measurement of presence in virtual reality. Presence is highly subjective, so there are still many debates among researchers about how to measure it, ranging from subjective measures (self-report questionnaires) to objective behavioral measures (reflex/startle response) to physiological measures (changes in heart rate, skin conductance, etc.) [10].

An open question is that behavioral and physiological measures could only be used to infer presence in limited circumstances, particularly for environments that cause measurable arousal (either with a negative or positive effect) [11,12]. In circumstances where no arousal effects are expected, this method cannot be used, or only where there are specific triggers exclusively to provoke physiological or brain responses. In this sense, there is a question regarding the possibility that immersive simulations of critical situations in height training or confined spaces would be capable of causing measurable arousal that would allow us to adopt measures such as heart rate variability or startle frequency, for example, to infer presence.

Virtual reality (VR) and augmented reality (AR), which have already been extensively used in immersive training [4,6,9,11], are promising for overcoming these practical and methodological challenges in two main ways. The first way is that the high degree of presence provided by immersive technologies can advance training evaluations toward Level 3 (behavior) of the Kirkpatrick Model. A successful VR experience makes people believe they are truly in a different environment [4,10–13]; a high sense of presence leads trainees to engage more with the training and be more likely to behave in the simulated environment as they would in the real world, as people who are afraid of heights in the real world will be afraid in VR [10]. It is possible, then, to immerse trainees in realistic simulations of their high-risk industrial processes (such as occupational risk situations in work at heights, fire, or confined spaces) and observe behaviors to assess whether they put into practice the knowledge acquired during training in their daily activities [8].

The second way in which immersive technologies can overcome the methodological challenges of training evaluation is through the measurement of physiological and behavioral data, which can be captured in simulations that record a trainee's actions [14]. Traditional assessment methods like interviews, questionnaires, knowledge tests, and practical tests are relatively easy to implement [12], but they are not as effective in measuring the development of complex skills (such as problem-solving, teamwork, and collaboration), and in the case of immersive training, they interrupt the simulation. These complex behaviors are especially relevant in high-risk industrial scenarios, where effective training must translate into adaptive decision-making and safe operational conduct. The present study addresses this by proposing evaluation guidelines focused on such behavioral dimensions.

A promising approach to capturing these behavioral and cognitive responses is through biofeedback, which is a method that uses biosensors, such as electrodes, to measure the patient's physiological reactions in real time. Various VR/AR head-mounted displays already integrate multiple physiological sensors to capture EMG—muscle movement; EOG—eye tracking; EEG—brain activity; EDA, PPG—heart rate; ECG; GSR; as well as facial and body tracking [13,15]. These physiological responses are not conscious processes, and, in general, can be identified using different measures, such as skin electrodermal activity and heart rate.

Thus, capturing these physiological and behavioral data generated by the trainee will enable the creation of more comprehensive, including automated, assessments that encompass complex skills, such as problem-solving, which translate into better real-world performance.

In light of these challenges and opportunities, this study presents preliminary design guidelines to support the evaluation of industrial safety training using immersive technologies, with a focus on high-risk work environments such as working at height. The findings aim to advance the understanding of how organizations can adopt immersive technologies—integrated with the capture of physiological and behavioral data—to assess performance and promote more effective training outcomes.

The paper is structured into three sections: Section 2 outlines the materials and methods adopted, Section 3 examines the findings, and Section 4 offers our final considerations and recommendations for future studies.

2. Materials and Methods

This study is exploratory because research on industrial training evaluation using immersive technology is recent, making qualitative research a more suitable approach [16]. Three senior researchers in virtual reality and industrial applications defined the initial research strategy, which was then assessed by five specialists and students, and finally, all authors peer-debriefed the strategy.

The design science research (DSR) paradigm was followed, which not only contributes to the advancement of knowledge but also addresses real-world applications related to the research problem or opportunity [17]. According to the DSR artifact taxonomy, our study provides an artifact of type design guidelines [18]. The following DSR's six steps were followed: (1) identify the problem, (2) define the solution objectives, (3) design and develop the artifact, (4) demonstrate it, (5) evaluate the artifact, and (6) communicate the results.

In DSR steps 1 and 2, the problem was identified, and the objectives of the industrial training evaluation using VR technology were set by conducting a systematic literature review on the subject, whose results were published [19].

In step 3, which involves the design and development of the artifact, we organized the literature review results into preliminary categories. We then conducted an exploratory focus group with eight interdisciplinary experts from our team, each one a professional with formal credentials, substantial practical experience, or recognized knowledge (referred to as Expert 1 to Expert8), listed below in Table 1.

The group meeting was conducted online, using Google Meet with audio and video; all participants could see each other and freely speak at all times. Access was managed to ensure privacy and freedom, and all participants gave consent to audio recording. Starting with the presentation of the topics by Expert 1, the meeting lasted approximately 80 min. The discussions were transcribed using the Read AI.

Based on our observations, we compiled the design guidelines. According to Design Science Research, these guidelines serve as support and provide suggestions for behavior in specific situations. A guideline offers a generalized suggestion for system development, often structured as statements like “in situation X, one could/should do Y” [18]. Qualitative research analyzes textual information, allowing researchers to interpret emerging themes or patterns from data [16]. We analyzed the data to extract meaning from the text, which involved segmenting, deconstructing, and reconstructing information [16]. Then, we organized the experts' discussion themes, and therefore the proposed guidelines, into the following six categories: scenario, procedures and ethics, recruitment, equipment, experimental design, and implementation. To improve the understanding of this process and our findings, we have created visual summaries and diagrams. Figures 1, 6 and 7 were produced by the authors, while Figures 2, 3, 4 and 5 were developed by the authors

using Napkin.ai as a design aid, and were subsequently post-processed to ensure visual consistency across the manuscript. Figure 1 below represents the DSR methodology applied in the current research.

Table 1. Participant backgrounds.

ID	Experience (Fields/Areas)	Experience (Years)	Gender	Age (Years)
Expert 1	Human-computer interaction, Immersive Technologies, Industrial Management.	17	F	48
Expert 2	Management (Industrial and Public), Organizational Innovation and Transformation, Worker Health and Safety.	15	M	50
Expert 3	Extended Reality (XR) and Metaverse applications, Game Design, and Development of Immersive Experiences for Education and Training.	7	M	29
Expert 4	Teaching and clinical practice in the field of career counseling and people development.	12	F	44
Expert 5	Innovation Management in construction sites and Human-Computer Interaction.	15	F	64
Expert 6	Senior Engineering professor and researcher.	15	M	42
Expert 7	Human-computer interaction, Extended Reality (XR), Tangible Interfaces, Immersive Learning, Virtual Humans.	20	M	45
Expert 8	Mechanical engineering, management and academic activities.	20	M	46

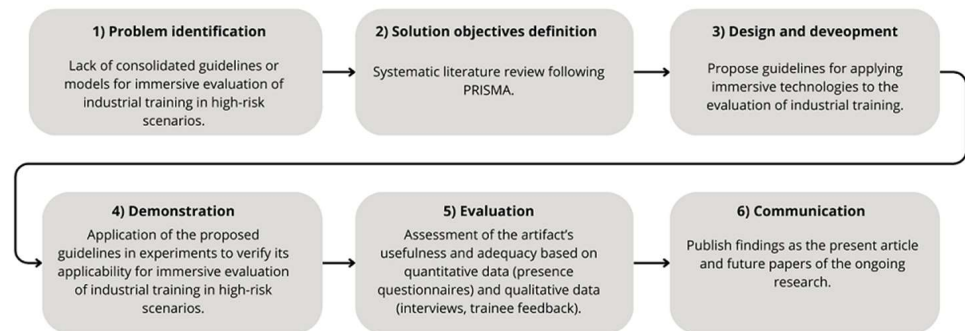


Figure 1. Design Science Research path, adapted by the authors.

Steps 4 and 5 of the Design Science Research (DSR) process—demonstration and evaluation—will consist of the development and validation of these proposed guidelines, in future proof-of-concepts. Preliminary results have been promising [15], and the research is currently ongoing.

The study was approved by the CIMATEC Ethics Committee (7.012.542). All research participants gave written and informed consent.

Step 6 of DSR consists of this publication to communicate the results of the design and development phase (DSR Step 3). In this phase, the expert focus group provides the foundation for future immersive evaluations aligned with Level 3 of the Kirkpatrick Model. The outcomes of this stage are detailed in the following sections.

3. Results and Discussion

In the following sections, we present and examine our experts’ discussions, as well as the proposed guidelines related to scenarios, procedures, ethics, recruitment, equipment, experimental design, and implementation.

The experts positively evaluated the focus group, highlighting aspects such as the diversity of experiences, backgrounds, and qualifications present. The interdisciplinary discussion covered several topics, which are visually summarized in Figure 2:

- Application of immersive technologies in industrial training.
- Presentations and discussions on ongoing projects and research.
- Application of virtual reality in industry and digital skills.
- Client requirements for training at heights.
- Simulation of wind and external stimuli for height training.
- Simulation of wind and other external stimuli in virtual training.
- Decisions and complexity in implementation.
- Tracking technologies and limitations of accessories.
- Discussion on the development of virtual reality training.
- Limitations of the technology for the project.
- Development of meaningful immersive experiences.












Topic	Description
 Industrial Training	Application of immersive technologies
 Ongoing Projects	Presentations and research discussions
 VR in Industry	Application and digital skills
 Training at Heights	Client requirements
 Wind Simulation	External stimuli for height training
 Virtual Training	Simulation of external stimuli
 Implementation	Decisions and complexity
 Tracking Tech	Limitations of accessories
 VR Training Dev	Discussion on development
 Tech Limitations	Project constraints
 Immersive Experience	Development of meaningful experiences

Figure 2. A summary of the key topics discussed during the focus group. This diagram was generated using the Napkin.ai tool, based on the textual descriptions of the discussion points.

From the aforementioned discussion, several action items and key questions were identified to guide future research and development in immersive industrial training. One of the primary actions is to evaluate the feasibility and effectiveness of using only virtual reality (VR) headsets for training purposes, without the need for additional hardware. Another important initiative involves researching and developing technologies capable of realistically simulating work-at-height conditions, which are critical for safety training

in industrial contexts. In this regard, the proposal to create a model that helps determine when it is beneficial to adopt immersive technologies in training becomes essential for strategic decision-making.

Additionally, it is necessary to analyze the feasibility of using advanced devices, such as the Pico 4 Enterprise, to collect physiological and stress-related data during training sessions. This exploration also includes the possibility of integrating biometric sensors into the safety harness typically used in height-related tasks, enabling data collection without interfering with the trainee’s experience. Another proposed action involves the development of a training station equipped with an elevating platform, which could replicate real working conditions and enhance immersion. Finally, there is a need to define and incorporate examples of unsafe scenarios that can be simulated in virtual environments to reinforce risk perception and safety behavior.

Several key questions also emerged during the discussion. For instance, to what extent is it beneficial to integrate external stimuli, such as wind or odors, into VR training to enhance realism and immersion? How can these elements be effectively simulated within a virtual environment? Further questions focus on the technical feasibility of transferring biometric sensors to wearable safety gear and how best to design the experience to distract users from the artificiality of the simulation, thereby increasing their sense of presence in the virtual environment.

This section presents and analyzes the results from the focus group with industry experts. The findings are organized according to a logical progression of key stages required for developing immersive training, from establishing the initial scenario to creating final design guidelines. This structure, which guides the following sub-sections, is visually presented in Figure 3.

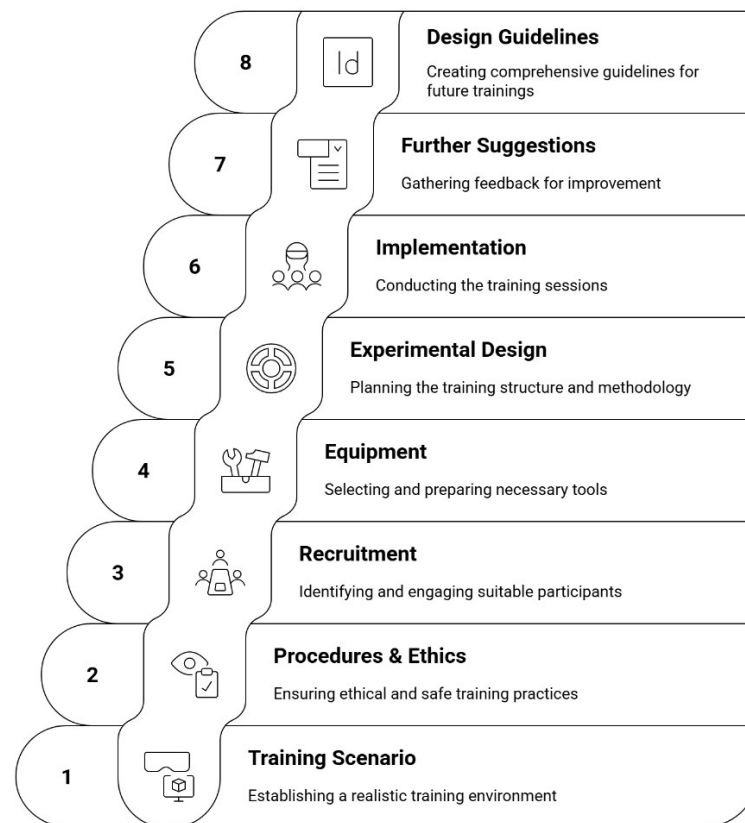


Figure 3. The structure of the results and discussion section, from the initial training scenario to the final design guidelines. This diagram was generated using Napkin.ai to visually represent the chapter’s workflow.

3.1. Training Scenario

The most appropriate scenario for the proof of concept was discussed, and height work training was recognized by the experts as having great potential for immersion with the current state of technology.

“And the height (. . .), because the stereoscopy that virtual reality gives, the fact of having two lenses, we have this spatial perception, makes us really feel that sense of height. So we can trick the brain to the point where the person understands that they are at a height of 3, 9, 30 m” (Expert 3).

Some experts mentioned their experiences with virtual training and the difficulties of evaluating whether it was effective. Thus, the research gap was reinforced, and the interest in filling a knowledge gap, as well as a need for companies, was emphasized.

“(. . .) the actual practice stage, right, and the validation of whether the operator learned in the theoretical part, whether they are indeed ready to practice it.” (Expert 3).

After identifying the typical mandatory training conducted in an industrial plant, such as Safety in Electrical Installations and Services, Transportation, Handling, Storage, and Material Handling, or Safety and Health at Work with Flammables and Combustibles, the group of experts assessed that it is feasible to develop and evaluate training that qualifies workers to understand and apply preventive measures, planning, organization, and execution of work at height. The choice is justified by the annual frequency of application, the related risk of accidents, the high subsequent replicability in various other industrial sectors, and the technological maturity of virtual reality that provides an adequate level of presence for immersive height experiences.

To decide what activities would be performed, Expert 5 presented a list of typical activities and the typical protection gear needed, summarized below:

In industry (general):

- Maintenance and replacement of light bulbs;
- Work performed in depth;
- Maintenance of electrical networks, towers, transmission lines, and antennas;
- Industrial cleaning and maintenance;
- Maintenance of furnaces and boilers;
- Assembly and disassembly of structures.

In Civil Construction:

- Maintenance and construction of roofs and coverings;
- Assembly and disassembly of shoring structures;
- Assembly of prefabricated structures;
- Concrete pouring of structures;
- Facade cladding;
- Maintenance and/or painting of facades;
- Installation of window frames.

Movement restriction equipment, like belts that keep workers from reaching the edge of a slab or risky location, protects them from the risk of falling. Fall arrest technology, such as safety nets or parachute-type safety belts, does not prevent falls but rather stops them after they have begun, lowering the effects. Depending on the height, the worker may also need to use a suspended work platform, suspended chair, or scaffolding to reach the position safely.

Considering the logistical challenge, combined with the safety precautions, the proof of concept will be developed in a controlled environment, preferably a laboratory with similar conditions. The experiment should take place on the SENAI/CIMATEC premises,

with an industrial rope access suspended from the ceiling and 1.0 m from the floor, so that the experiment participant's legs are in the air. And at IFBA (Salvador Campus), in the testing laboratory, a platform will be constructed 20 cm from the floor and coupled to an existing iron structure, providing a balance comparable to that of a platform used for painting/coating services on façades on building sites.

Before the experiment begins, a brief presentation will be given to the study participants, who will be requested to complete a free permission form.

To begin the experiment, the student will stand on the platform or sit in the individual rocker and put on the virtual reality goggles. From that point on, the researcher will observe their movements and collect data using the sensors put in the VR goggles and the watch on their left arm. This experience should take around 15 min.

After wearing the head-mounted display and experiencing the immersive evaluation, the subject will complete a questionnaire to record their impressions about the experiment.

To conclude this discussion, the focus group discussed several industrial training contexts, ultimately selecting “working at height” due to its high accident risk, broad applicability across sectors, and strong alignment with VR capabilities such as stereoscopic depth and spatial immersion. While other scenarios like electrical safety or material handling were mentioned, working at height offers an optimal balance between technical feasibility and training impact, enabling realistic simulations that can effectively assess risk perception and behavioral responses in elevated environments. This scenario selection is a pivotal stage within our overall research workflow, which is visually detailed in Figure 4.

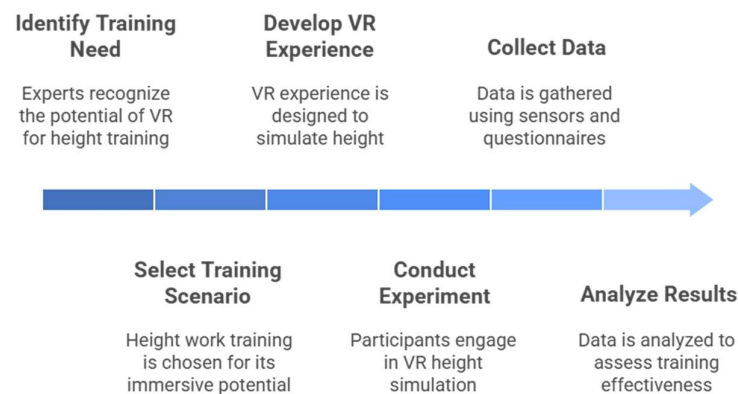


Figure 4. A visual representation of the research process, from identifying the training need to analyzing the results. The workflow diagram was created using the Napkin.ai tool.

3.2. Procedures and Ethics

Ethical concerns were discussed due to it being an experimental study, and the necessary mitigation and control measures to proceed were considered. The importance of having healthy and well-informed participants with no history of trauma from falls from heights was emphasized.

There are minimal risks related to the use of VR equipment, particularly the headset. Some individuals may experience dizziness, for instance, in addition to an emotional reaction to the simulation stimuli. Expert 4 stated, “I was thinking about the subject’s history and how much this history can impact the training moment”, aligning with Expert 1’s concerns. Therefore, to mitigate any health risks to the participants, an anamnesis will be conducted by a legally qualified health professional, including, among other items, a review of any history of accidents related to working at height or any other related major psychosocial risk factors.

The benefits of the research have the potential to reduce risks and accidents, justifying the exploratory study to contribute to the mitigation or even reduction of the high incidence of incidents and accidents.

“Because the risk for him to actually climb the scaffold is already assessed and managed by a competent, qualified professional. We are, in fact, bringing it to the immersive experience, which brings some other risks, but we have already taken an important step. The issue of benefits, because if we can indeed (. . .) mitigate the problems our country and the world have in relation to this, we have already made our contribution” (Expert 2).

It was also clear that there is potential to contribute to work regulation, especially in improving regulatory standards for work by possibly requiring virtual experience before exposure to real risks.

There are already companies in various parts of the World offering virtual reality training services, and it is necessary to build knowledge about their application to maximize positive impacts and avoid negative externalities. All participants will be informed about the experiment, with an emphasis on what virtual reality is, the purpose of the research, and clarifications that participation is entirely voluntary and that the data will be anonymized for publication. All will receive informed consent forms to be signed, and all possible questions will be answered.

In this context, a virtual reality experience will be developed that simulates working at height in two conditions: normal operation and emergency situation. Based on the results of the literature review [19], the following requirements were proposed:

The proposed environment for the proof of concept will incorporate realistic height variations—specifically 3, 9, and 30 m—to simulate approximate elevations corresponding to one, three, and ten stories, respectively. One of these options will be selected for implementation in the prototype. The simulation will be set in open or semi-open environments to enhance the perception of height. An avatar with average human proportions will be used, allowing for natural interactions involving arm and hand movements, as well as engagement with objects. Vertical movement will be enabled via stairs or scaffolding, operated through VR controllers, while horizontal movement will rely on the user’s physical body. The scene will include both fixed elements, such as structural installations, and dynamic components, such as birds, to enrich immersion. Users will also interact with tools and interface elements like buttons and levers. Simulated gravity will affect both falling objects and avatars, contributing to realism. Finally, instructional content will be available within the environment in the form of text and/or video, accessible as needed by the participant.

In summary, the group identified important ethical considerations to be addressed to avoid or mitigate any impacts of the experiment, such as, approval from the ethics committee; healthy workers (with occupational health certificates to work in the industry); the benefits of the research have the potential to reduce risks and accidents; the current existence of companies offering similar services; worker health monitoring (nursing and medical teams on standby to provide assistance if necessary); and the need to build useful and applicable knowledge, as guided by DSR.

Exclusion criteria for participants with a history of severe mental health issues, such as previous or current diagnosis of anxiety disorders, depression, schizophrenia spectrum disorders, mood disorders, or any psychological condition that, at the discretion of the researcher, may compromise the participant’s ability to understand and complete the study procedures safely and effectively, aiming to ensure the safety of participants and the integrity of the data collected.

Additionally, a history of serious accidents involving falling from heights, traumatic events related to heights, uncontrolled hypertension, or severe dizziness when using virtual reality.

In summary, the ethical framework proposed by the experts emphasizes proactive risk mitigation, participant well-being, and methodological rigor. This includes exclusion criteria based on trauma history and mental health, medical screening through anamnesis, and transparent informed consent. These measures not only protect participants but also support the scientific and regulatory credibility of immersive training applications. A summary of these key considerations is presented in Figure 5.





Characteristic	Participants	VR Experience	Research Benefits
 Ethical Review	Approval from ethics committee	Informed consent emphasized to all participants	Potential to improve work regulation standards
 Participant Health	Healthy workers with occupational health certificates	Anamnesis by health professional to mitigate risks	Potential to reduce risks and accidents
 Risk Mitigation	Exclusion criteria for mental health issues	Minimum risks related to VR equipment use	Companies offering similar services already exist
 Safety Measures	Exclusion criteria for height-related trauma	Nursing and medical teams on standby	Need to build useful and applicable knowledge

Figure 5. Summary of the ethical framework and risk mitigation measures discussed by the experts. The diagram was created using the Napkin.ai tool to visually organize the key considerations.

3.3. Recruitment

The proposed design consists of 30 participants, divided into 3 groups, selected in an intentional sample (where participants are chosen for having relevant characteristics), with the following profiles:

- Group 1: Students from construction, engineering, technology and related areas.
- Group 2: Construction, engineering, safety or industrial workers.
- Group 3: Engineering, Virtual Reality, Health and Safety, Training, Industrial Management and other professionals.

An anamnesis will be conducted with all participants in groups 1, 2, and 3 to preserve mental health and control variables, according to the exclusion criteria mentioned in the previous chapter.

To sum up, the recruitment strategy involves intentional sampling across three participant groups, with careful screening to control individual differences. The inclusion of students, professionals, and workers ensures diverse perspectives, while the use of anamnesis seeks to eliminate potential biases linked to prior trauma or health conditions, reinforcing the robustness of the experimental design.

3.4. Equipment

Alternatives for developing solutions to bring more realism to the participant's experience were discussed, including various technologies such as devices (headsets, controllers, sensors, etc.) and installations (stairs, platforms, fans, etc.). The discussion aligns with findings from recent studies on immersive technology platforms for training, which are limited by the available hardware and software. The degree of realism and immersion depends, among other factors, on the advancement of these technologies combined with the level of investment by the company [15,19,20].

Thus, proposals and limitations emerged during the discussion on ways to simulate working conditions at height, such as strong winds, as described by a specialist.

“He (the worker) was descending on the rope, that swing, that little chair, right? The chair with the swing (bosun’s chair). And after he started doing maintenance, it started to get windy. And then he had to wait two hours for the wind to reduce so he could descend. So it was a moment like, two hours. We think two hours is a long time for a person to stay hanging there” (Expert 5).

This discussion adds to the concern about the decision-making carried out by the worker as working conditions present themselves or change during height operations.

“Can we simulate wind? Because we know that NR35 states that if you have wind over 12 km per hour, or 45 km, no, 45 km per hour, or 12 m per second, you shouldn’t work. The ideal is between 8 and 10 m per second” (Expert 5).

Ways to simulate winds in virtual reality with the available technologies were presented, ranging from the most basic, such as using sounds, to the most elaborate, which would include fans with air speed measurement and monitoring.

“We have two ways, the first with sound. The first one (. . .), we can simulate wind sounds, and then we can even test to see if just sound can give the same sensation. (In the second way) we can put a fan next to the person or something that we can use to stimulate the same amount of real wind, you know. I don’t know if there are devices that we can control the wind speed from” (Expert 3).

Regarding the number of devices for the experiment, considering the possibilities of interference in the experiment, since the workers do not use them on their bodies during their daily routine, proposals were made to develop technologies, especially a safety belt with sensors already attached, which would be more natural for the participant to use.

“Couldn’t we transfer all the sensors to this protective gear, this safety belt, and keep it in contact with the skin? I don’t know how it works, but that’s what I thought, because there’s no work at height without this belt (. . .). So, instead of having a bracelet, this and that, you would wear this belt that fastens here on the thigh, works like a backpack, and has a hook on the back. Then we would try to put (. . .). He would just wear this belt, and that’s it, it would be gathering all the information. I don’t know if that’s possible, people? I’m just giving this suggestion” (Expert 5).

The suggestion to centralize the sensors on the safety belt was appreciated by the group, and favorable arguments emerged, as it would greatly reduce the interferences in immersion that could be caused by the excess of devices. However, limitations were presented that this restriction would bring, especially regarding capturing the participant’s body movements with the currently available technology, as exemplified in the following transcription.

“Taking out the belt would limit body motion, we wouldn’t be able to get the body’s movement itself, but limiting the number of accessories seems to be an interesting principle as well. If we could have what we need to measure biometric data in the mandatory accessories, it seems interesting to me” (Expert 6).

The specialists also considered the possibility of reducing the perception of multiple devices by combining them, as in the example below.

“The glasses themselves, both close to the temple area, maybe I can put a sensor in contact with the glasses’ environment? Maybe it’s better, I don’t know” (Expert 6).

The group considered that placing sensors under the headset might not be the best solution due to potential discomfort and the risk of incorrect readings. At the same time, a headset capable of heart rate measurement for stress identification was also considered; the

most suitable solution at the moment was determined to be the Samsung Galaxy Watch 7 (Figure 6), an established and readily available technology.



Figure 6. Samsung Galaxy Watch 7 for collecting heart rate data.

The Samsung Galaxy Watch 7, released in 2024, provides advanced health monitoring capabilities suitable for research purposes [21]. The device tracks heart rate and offers notifications for high, low, and irregular heart rhythms, suggesting a robust platform for cardiovascular monitoring.

“And then the other idea from Expert 6, who gave this great idea before, which is heart rate, I talked to a psychiatrist recently about another issue, and he said, look, if you can bring heart rate to VR, I can already read a lot, and stress is one of them, and stress we can bring to correlate with the immersive environment, how the person is, how the person is reacting, the behavior” (Expert 1).

“Great point, Expert 6. As a developer, I love challenges. So, for me, wow, put everything. Put clothes, platform, wind, I love it. For me, it would be incredible to develop. But I agree with Expert 2. What is the initial purpose? The initial purpose is to see if it is really worth developing this training in NR35, if it adds value within the limitation of the glasses” (Expert 3).

There was a brief discussion about the use of hands by the workers and how it would be interesting for them to use their own hand movements captured by the headset, instead of controllers. This would also avoid interference in the use of the watch to measure heart rate.

Finally, participants agreed that, for the present research, the technology to be used should be a headset capable of eye tracking, measurement of some vital signs, especially heart rate, and some form of body tracking, as long as it does not hinder the worker’s movements.

The group of experts recommended that the experience utilize the Pico 4 Enterprise virtual reality device (Figure 7), which has the eye-tracking feature incorporated. Models like the standard Pico 4 and the Meta Quest 3 were discarded as they do not have the eye-tracking functionality.



Figure 7. Pico 4 Enterprise is a Virtual reality head-mounted display.

The Pico 4 Enterprise, released in 2022, is a virtual reality headset featuring a high-resolution 4K+ display, a Pancake optical lens that affords a wide 105° field of view, and 6 Degrees of Freedom (6DoF) spatial positioning [22].

Overall, the discussion on equipment centered on balancing realism, usability, and technological feasibility. The recommendation to use integrated systems like smartwatches and eye-tracking HMDs—while exploring wearable sensor integration into safety harnesses—reflects a user-centered approach. Simulating environmental variables like wind through auditory and physical cues further enhances immersion while keeping the setup manageable and scalable.

3.5. Experimental Design

It is necessary to define how participants will be monitored, and for this purpose, the three strategies of presence used by instructors for virtual reality training were considered: having their own avatar to interact with the trainee; visualizing the behavior in real-time to guide the trainee; or post-experience analysis of the trainee through data collected by devices [8,14,19]. The first strategy involves co-presence for interaction with another avatar, which is not the object of this investigation stage. Therefore, a combination of the second strategy, in which researchers will monitor the participant in real-time, with the third strategy is considered more appropriate, as data will be collected through sensors for later analysis in complement to the participants' perception.

The theme “decision-making” emerged during the discussions, particularly about the importance of the decision not to carry out the activity, adapt it, or even interrupt it, as exemplified in the transcription below.

“We can create a simulation for him to make a decision. It could be in the training for him to decide whether to go up or not to work. If the wind is there or if he is working and the wind changes. Do you understand what I mean? If he decides to descend, I'm already imagining a scenario. I am already here in the scenario” (Expert 5).

This decision-making component reflects the broader problem-solving demands of high-risk tasks. The simulation aims to observe behavioral responses to changing conditions, such as wind intensity, to evaluate how participants adapt and make safety-critical decisions.

Regarding the strategy of adopting technologies in the experiment, questions also arose that apply to the design of effectiveness evaluation for immersive training and the decision on which devices and installations should be part of it. One of the specialists conducted a reverse reasoning, starting from the real need of the client and the possibilities of what can be performed with the available technology, transcribed below.

Specialists expressed concerns about the practical implementation of these ideas in the short term, and there was a suggestion to start with a more objective experience, using glasses with eye tracking and heart rate capture, before moving on to more complex stimuli.

“The glasses with eye tracking, that's already built-in, but the data isn't that easy to analyze, so that's a challenge. The ring or something else to capture heart rate, and the body tracker, and place the person on a virtual platform, put the person on the edge of a building, in short, an easier experience to carry out. And then we can add to it.” (Expert 1).

“Regarding what was being said about only using glasses, we had already talked a bit about this, I have often started to diverge, and you say, it's not just the glasses. And here the question is, if we only have glasses, I think it's very important, even from a pragmatic point of view also for Expert 3, it's very important to understand the opposite, which is, if we only have glasses, how far can we go?” (Expert 6)

The discussion continued with the reflection on the decision to adopt preparatory and planning measures to maximize immersion within the available technological limitations,

which would likely be more than sufficient for this stage of the research, considering the height work simulation and would open up for future studies.

Expert 6 pointed out that although high-fidelity simulators—such as cockpit or machine simulators equipped with force feedback and six-axis motion platforms—can deliver highly immersive experiences, they require workers to be physically present in specific facilities. While such setups offer significantly superior realism compared to standalone head-mounted displays (HMDs), they present logistical challenges. He emphasized the importance of investigating the potential of more accessible solutions, particularly those based solely on HMDs, and questioned the extent to which meaningful immersive experiences can be developed within these constraints.

According to Expert 6, understanding the limitations and possibilities of HMD-based simulations is critical, especially when considering client requirements for specific applications like work-at-height training. He recalled an early experience using VR goggles, where simply walking toward the edge of a skyscraper—despite the absence of wind or other sensory cues—successfully triggered a strong perception of height. This demonstrated that even minimal setups can evoke essential sensations. However, he argued that simulating realistic working conditions requires more than just visual immersion, raising the question of how far one can push such technologies without additional sensory input.

He further noted a psychological dimension: when users are aware they are in a virtual environment, they may consciously or unconsciously suppress emotional responses such as fear, knowing they are safe. This awareness could limit the effectiveness of the simulation. Therefore, the expert highlighted the relevance of integrating biometric data—such as body tracking, heart rate monitoring, and eye movement analysis—to help reduce this cognitive barrier and fully engage participants in the virtual domain. Ultimately, he reiterated the key question: how far can we go with immersive training when relying solely on head-mounted displays?

“Therefore, lowering the barriers to make the experience effective. And, once again, I say, if it is only with glasses, how far can we go with this situation?” (Expert 6).

Following the discussion above, the experiment will be conducted with the best physical conditions available to approximate the working situations, considering the resources available. There will be two possible scenarios: the first using wood to represent the simulated working platforms, and the second one, a bosun chair with additional safety gear will be used, at a minimum height (approximately 0.25 m).

The experience with the two training conditions in virtual reality (normal operation and emergency situation) will be created in the Unreal Engine/Unity game engine.

The initial proposal for external evaluation of immersive training was designed to involve three methods: individual assessment based on the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) [23,24], the Presence Questionnaire (PQ) [25,26], and interviews. The UTAUT2 model is widely used in research to assess technology adoption, while the PQ measures perceived presence. However, upon further analysis, concerns arose about the limitations of questionnaires and interviews. Although simple, they impose conceptual frameworks that may not align with the actual experiences of participants. Physiological and behavioral measures may offer a more objective approach [27], but their application is limited to environments where specific events trigger automatic responses, ensuring their reliability in detecting presence. Sentiment analysis provides quantitative and qualitative insights into participants' reactions, as demonstrated by Archer et al. (2022) [28], who found correlations between questionnaire responses and physiological measures in assessing presence related to virtual environments with odors.

Considering these factors, the ideal measurement should integrate multiple approaches, including setting transitions, qualitative methods such as sentiment analysis, and supporting survey results with behavioral or physiological data, where appropriate.

Sentiment analysis uses machine learning techniques to assess sentiment in texts using a pre-trained database. Essays analyzed using this method revealed clusters of sentiments, allowing researchers to identify common themes, such as realism, where higher sentiment scores correlated with descriptions of the virtual environment as highly realistic [20]. This method allows participants to reflect on their experiences in an unconstrained way, while also producing quantifiable data.

In addition to sentiment analysis, biofeedback techniques—including eye tracking, body tracking, and heart rate variability—can further enhance assessment. Experts recommend collecting biometric data, such as pupil dilation, fixation movements, blink rate, and heart rate variability, to improve understanding of participant behavior in immersive training. Traditional assessments rely heavily on self-reported experiences, which can miss gaps due to discomfort or forgetfulness. Tavares et al. (2024) [29] highlighted the potential of eye tracking to complement qualitative data and provide objective indicators of mental workload. Although eye-tracking-based metrics show promise for assessing patterns of stress, attention, and relaxation, more research is needed to refine their applications. Physiological responses have already been used as indicators of presence, as demonstrated by studies involving virtual environments [30]. For example, heart rate indicated arousal in situations such as witnessing a fire in a virtual environment [31,32] or how a stressful VR elevator produces physiological as well as psychological stress responses [33]. Brain activation has also been measured with EEG during VR studies [34]. Further investigation into the connection between physiological responses and presence measurement supports the use of complementary approaches for comprehensive assessment [32,34,35]. These physiological indicators will serve as complementary evidence of engagement, cognitive load, and behavioral adjustment in response to simulated challenges.

Therefore, presence measurement benefits from triangulation across multiple methods, ensuring both subjective and objective assessments. By integrating sentiment analysis, physiological responses, and behavioral observations, researchers can establish robust criteria for evaluating the effectiveness of immersive training and optimizing design tradeoffs [20].

Heart rate variability will be collected by devices such as electrocardiogram (ECG or EKG) systems, eye tracking will be collected by the sensors embedded in the head-mounted display, and for body tracking, sensors like the Movella XSense can be adopted.

To conclude this section, the experimental design reflects a pragmatic and scalable approach, combining real-time monitoring with post-simulation analysis. Emphasis was placed on simulating decision-making scenarios involving dynamic risk factors. The group also acknowledged the limitations of HMD-only setups, advocating for maximizing immersion within available technological constraints as a foundation for future refinement.

3.6. Implementation

Expert 3 emphasizes the importance of not replacing all traditional training with the use of immersive technologies.

“(.. .) virtual reality is very difficult for us to replace training. I think the standard training that exists in industries today (.. .) will continue to exist. I think virtual reality will be one of the stages of this training. I wanted to bring this up because I think we need to take this into account because it is very difficult to do 100% of the training in virtual reality” (Expert 3).

This understanding aligns with the findings in the literature on the category “Combined Real and Virtual Experiences,” in which it is concluded that the integration of immersive technologies into training programs is not intended to replace hands-on sessions with real equipment, but rather to complement them by combining physical and virtual experiences in a synergistic manner [8,11,19].

One point highlighted by experts is the perception of risk by workers, which may not align with factual reality. There are situations where the worker underestimates or overestimates their abilities to deal with adversities, which can lead to insecurity on one side or accidents on the other. This concern aligns with one of the individual characteristics highlighted in recent studies, called self-efficacy, which is the person's perception of their ability to learn and perform actions [4,19].

One of the experts summarized much of the discussion as follows:

"The synthesis of what we seek to understand (at this stage of the study) is the decision-making process of continuing or interrupting the activity when conditions change, especially if unsafe conditions arise (e.g., wind, lighting, rain, vibrations, instability in the scaffold, etc.)" (Expert 2).

Thus, there was a consensus on the positive potential contribution of using immersive technologies in training evaluations in work-at-height environments, provided there are guidelines for their application, which is in line with the gap in the present research.

In short, the implementation phase should integrate immersive technologies as complementary tools rather than replacements for conventional training. The emphasis on gradual transitions into and out of VR, along with real-time feedback and post-session reflection, aims to foster safe engagement, reinforce learning, and support the development of critical decision-making skills in high-risk contexts.

3.7. Further Suggestions

This section contains ideas and insights of high value that go beyond the scope of the current state of the research, such as themes for future studies.

Gamification was a theme brought up by two experts and corroborated by two more, due to the need to combine it with decision-making processes. *"One of the things I want to try to bring into the discussion is how we can gamify the training"* (Expert 3).

This is a rich area for research exploration, combining strategies for using virtual reality techniques, including missions, scoring, rewards, and others, seeking to understand how these influence decision-making processes and the future possibilities of applying emerging technologies.

There was also a proposal to study the real need for other equipment to simulate strong winds, which could be used to simulate the decision-making process of stopping a risky activity, but which would depend on studies about the effectiveness of each alternative, as highlighted below.

"Mixing sound with the physical stimulus, we can do it without problems. It's even nice that we can make a comparison in the research. Sun and sound, just the wind, wind and sound, anyway. I loved this idea; I think we could bring this into the experience" (Expert 3).

A recurring theme is workers' resistance to using protective equipment, which can be analyzed from various perspectives (legal, motivational, decision-making, ergonomics) and could be explored with the use of immersive technologies in future studies.

Future research should prioritize the integration of gamification elements into immersive training environments, particularly focusing on how missions, scoring systems, feedback loops, and reward structures influence decision-making under risk. Experimental designs employing comparisons can reveal how different gamified strategies impact engagement and behavioral outcomes [36,37]. This line of inquiry has the potential to generate technological innovations in training platforms, while also promoting cost-effective and scalable safety interventions. Socially, it may help foster a culture of proactive behavior and shared responsibility for workplace safety, especially in industries with historically high accident rates.

Another promising avenue involves simulating adverse environmental conditions, such as strong winds, to explore their effects on the user's risk perception and decision-making about interrupting unsafe tasks. Mixed-method studies combining performance metrics and qualitative feedback can assess the relative effectiveness of multisensory stimuli (audio, visual, tactile) in creating realistic yet safe simulations. The resulting knowledge can inform the design of more adaptive and resilient training systems, with economic benefits by reducing the need for costly real-world replications and enhancing workforce preparedness for unpredictable conditions.

Additionally, research should investigate worker resistance to protective equipment through interdisciplinary frameworks—incorporating ergonomics, behavioral psychology, motivation, and legal perspectives. Immersive simulations offer a controlled environment to test interventions meant to boost compliance and shift safety-related attitudes. Such findings could directly contribute to reducing accident rates, lowering insurance and compensation costs, and improving long-term health outcomes for vulnerable or precarious workers, aligning with global agendas such as the United Nations' Sustainable Development Goals [1].

Moreover, further research is needed to develop and validate models for assessing the effectiveness of safety training—particularly in immersive and virtual environments. Traditional evaluation methods often rely on declarative knowledge tests or self-report measures, but immersive technologies open new opportunities for dynamic, behavior-based assessments. Carnell et al. (2022) demonstrated how the Kirkpatrick Model can be applied within a virtual human scenario to evaluate behavioral changes in communication skills, showing the potential of VR to support structured, multilevel evaluation frameworks [38]. In parallel, recent research has explored the use of physiological and behavioral biofeedback to assess user experience in immersive safety training, capturing stress, cognitive load, and attentional shifts in real time [39]. Additional studies have proposed integrated approaches combining presence, usability, and learning outcomes to better capture the complexity of immersive learning [40], while others emphasize the value of real-time behavioral analytics and eye-tracking to measure decision accuracy under risk [41]. In a complementary direction, Doolani et al. (2023) introduced a comprehensive framework for evaluating XR-based safety training, integrating physiological data, task performance, and long-term transfer metrics [42]. By combining immersive analytics with validated evaluation models, these approaches lay the groundwork for designing adaptive, personalized, and evidence-driven training systems—especially critical in high-risk occupational contexts.

Finally, extended reality (XR) solutions—encompassing VR, AR, and MR—should be analyzed in terms of their broader impacts on occupational safety and innovation in high-risk industries. There are opportunities to improve safety knowledge, productivity, and instructional delivery while recognizing technological and human-factor challenges. Building on these findings may guide the development of more effective, personalized, and context-sensitive training models, ultimately strengthening institutional safety cultures, enabling cross-sector digital transformation, and supporting evidence-based policy development.

Looking ahead, the experts identified valuable avenues for future exploration, including gamification to support engagement and decision-making, simulation of adverse conditions, and analysis of worker resistance to safety practices. These insights suggest a rich agenda for advancing immersive training through interdisciplinary research and evidence-based policy contributions.

3.8. Design Guidelines for Immersive Industrial Training Evaluation

We compiled the experts' discussion themes about scenario, procedures and ethics, recruitment, equipment, experimental design, and implementation into six preliminary guide-

lines categories. They are phrased as generalized suggestions regarding behavior in specific situations, and formatted as statements, following the Design Science Research principles regarding guidelines artifacts [17,18]. Table 2 below presents our proposed guidelines.

Table 2. Immersive industrial training evaluation preliminary guidelines.

	In Situation XOne Should Do Y
Scenario preliminary Guideline	When setting the scenario . . .	Analyze the main occupational risks present in the field of study by degree of risk, frequency or severity.
		Select one that can be adequately simulated with the available technology.
		Seek to maximize the effectiveness of immersive assessment within technological limits.
		Define which safe behaviors should be evaluated with the aim of mitigating or eliminating risks.
		Define which behaviors should be monitored during the immersive assessment.
Procedures and ethics preliminary Guideline	When evaluating ethical procedures . . .	Identify the working conditions under which the actual work is performed.
		Mitigate risks to the participant.
		Present all the conditions for an immersive experience.
		Check possible psychological risks and whether the participant has any traumatic experiences related to the object of the immersive assessment.
		Check whether the participant has ever had severe symptoms from using VR.
Recruitment preliminary Guideline	When selecting participants . . .	Check if participants have previous training.
		Check if participants have previous hands-on experience.
		Conduct an anamnesis in advance to check for a history of serious accidents involving working at heights.
Equipment preliminary Guideline	When defining devices and facilities . . .	Evaluate what is possible to simulate, within the available technology, and align the expected results.
		Include previously identified working conditions.
		Choose the most appropriate equipment for the selected tasks and tools.
Experimental design preliminary Guideline	When designing the immersive assessment . . .	Evaluate what is possible to simulate, within the available technology, and align the expected results accordingly.
		Use elements of real work associated with the content to be evaluated.
		Identify which skills will be prioritized for assessment.
		Define the level of realism versus playfulness.
		Consider gamification elements related to the skills to be assessed to foster engagement.
Implementation preliminary Guideline	When implementing . . .	Ensure that the start of the immersive assessment is suitable for immersion, and the exit, back into the “real world”, is also gradual.
		In the event of detection of a simulated incident or accident, validate and support the participant.
		Record the experience, successes and errors, with real-time feedback to the participant (sounds, texts).
		Provide learning opportunities to participants.
		Use the results of the immersive experience to improve training programs.

In accordance with the Design Science Research (DSR) paradigm, the guidelines developed in this study represent a prescriptive artifact that must undergo rigorous testing and iterative refinement to demonstrate utility and validity in real or simulated contexts [17]. Following the completion of this conceptual design phase (DSR Step 3), the next step (DSR Step 4) will involve the implementation and demonstration of the artifact in a controlled environment. To this end, the guidelines will be applied in the design of an immersive training evaluation prototype, focusing on occupational safety in high-risk environments.

This prototype will be tested in a laboratory setting through a realistic simulation involving virtual reality technologies, interaction tracking, and performance monitoring for work at height. The aim is to assess not only the feasibility of implementing the proposed guidelines but also their practical effectiveness in guiding immersive evaluation processes. This approach aligns with the DSR methodology, which emphasizes the importance of empirical demonstration to ensure that the designed artifact effectively addresses a real-world problem while also contributing to the body of scientific knowledge.

Taken together, the proposed set of six guideline categories constitutes a theoretically grounded and practically oriented framework for the evaluation of immersive industrial training. Derived through interdisciplinary expert analysis and structured according to the Design Science Research paradigm, these guidelines serve as a prescriptive artifact aimed at supporting the rigorous design, implementation, and assessment of VR-based training interventions in high-risk occupational contexts. As such, they offer a foundation for future empirical validation and iterative refinement in both academic and applied settings.

To ensure transparency and scientific rigor, several limitations of this study must be acknowledged. First, the number of participants was limited, and the sample was composed exclusively of subject-matter experts. Although this selection aligns with the exploratory nature of the study, it restricts the generalizability of the findings. The perspectives of frontline workers, operational staff, and stakeholders from diverse cultural or industrial contexts were not represented in this initial stage. Including these perspectives in future iterations may help enhance contextual relevance and broaden the applicability of the guidelines across different training environments.

Moreover, as is common in qualitative research, the group dynamics and the facilitation process may have influenced the depth and distribution of individual contributions. The analysis and categorization of thematic content were conducted with methodological rigor; however, the findings are still subject to interpretative bias and contextual influence, particularly given the specificity of the Brazilian industrial training landscape.

It is also important to emphasize that this study represents an intermediate stage within a broader Design Science Research (DSR) cycle. The proposed guidelines constitute an early iteration of the artifact and have not yet undergone empirical validation in real-world training environments. Future research phases will involve applying the guidelines to the development and implementation of immersive evaluation systems for industrial workers, with a particular focus on assessing behavioral change in high-risk training scenarios, in line with Level 3 of the Kirkpatrick Model. These subsequent stages will be essential to assess the artifact's effectiveness, scalability, and practical relevance, and to iteratively refine the model based on empirical evidence. Complementary studies may also incorporate quantitative data and multidisciplinary feedback to further enhance the robustness and transferability of the proposed framework.

4. Conclusions

This study contributes to advancing the use of immersive virtual reality technologies in the evaluation of industrial training, particularly in high-risk scenarios such as working at height, and lays the groundwork for future behavior-focused assessments aligned with Level 3 of the Kirkpatrick Model. Through the application of the design science research methodology and engaging interdisciplinary experts, we identified six key categories and formulated preliminary guidelines to support the design and implementation of immersive training evaluations. These guidelines offer a structured approach to simulate realistic work conditions, monitor participant behavior, and collect objective biometric data to assess learning transfer to actual professional behavior. The findings underscore the importance of combining subjective and physiological measures to enhance the validity of

presence and behavior assessments in virtual environments. Although further development and empirical validation of the proposed guidelines are necessary, this study lays the groundwork for designing more effective, safe, and expert-informed training evaluation systems in industrial settings.

Author Contributions: Conceptualization, A.C.; methodology, A.C. and I.W.; validation, R.L., L.A., C.N., M.C., T.S. and A.S.; formal analysis, A.C.; investigation, A.C., R.L., L.A., C.N., M.C., T.S., A.S. and I.W.; writing—original draft preparation, A.C., I.W. and R.L.; writing—review and editing, A.C., I.W. and R.L.; supervision, I.W., A.S. and T.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Advanced Knowledge Center in Immersive Technologies (AKCIT), the Bahia State Research Support Foundation (FAPESB), and the National Council for Scientific and Technological Development; grant number 308783/2020-4 and the APC was funded by the Advanced Knowledge Center in Immersive Technologies (AKCIT).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the CIMATEC Ethics Committee (code 7.012.542, 17 August 2024).

Informed Consent Statement: All participants provided written informed consent.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: The authors thank the SENAI CIMATEC University, Oswaldo Cruz Foundation. The authors acknowledge the use of Napkin.ai for generating diagrams supporting the development of Figures 2–5.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. International Labour Organization (ILO). *Relevant SDG Targets Related to Health and Safety at the Workplace*; International Labour Organization: Geneva, Switzerland, 2023.
2. Tribunal Superior do Trabalho (TST). *Acidentes de Trabalho Matam ao Menos Uma Pessoa a Cada 3 h 47 min No Brasil*; Tribunal Superior do Trabalho: Brasilia, Brazil, 2022.
3. SmartLab. Observatório de Segurança e Saúde no Trabalho. 2025. Available online: <https://smartlabbr.org/> (accessed on 3 July 2025).
4. Chen, C.; Ping, S.; Zhang, X.; Yi, Y. Transfer study of safety training based on mapping knowledge domain—Overview, factors and future. *Saf. Sci.* **2022**, *148*, 105678. [[CrossRef](#)]
5. Kirkpatrick, D.L.; Kirkpatrick, W.K. *Kirkpatrick's Four Levels of Training Evaluation*; Association for Talent Development: Alexandria, VA, USA, 2016.
6. Zhang, M.; Hou, L. Formulating a learner model for evaluating construction workers' learning ability during safety training. *Saf. Sci.* **2019**, *116*, 97–107. [[CrossRef](#)]
7. Kirkpatrick, D.L.; Kirkpatrick, J.D. *Evaluating Training Programs*; Tata McGraw-Hill: Columbus, OH, USA, 2006.
8. Perez-Ramirez, M.; Arroyo-Figueroa, G.; Ayala, A. The use of a virtual reality training system to improve technical skill in the maintenance of live-line power distribution networks. *Interact. Learn. Environ.* **2021**, *29*, 527–544. [[CrossRef](#)]
9. Abich, I.V.J.; Parker, J.; Murphy, J.S.; Eudy, M. A review of the evidence for the effectiveness of training with virtual reality technology. *Virtual Real.* **2021**, *25*, 919–933. [[CrossRef](#)]
10. Jerald, J. *The VR Book: Human-Centered Design for Virtual Reality*; ACM & Morgan & Claypool: San Rafael, CA, USA, 2015.
11. Freina, L.; Ott, M. A literature review on immersive virtual reality in education: State of the art and perspectives. *Elearning Softw. Educ.* **2015**, *1*, 133–141.
12. Makransky, G.; Klingenberg, S. Virtual reality enhances safety training in the maritime industry: An organizational training experiment with a non-WEIRD sample. *J. Comput. Assist. Learn.* **2022**, *38*, 1127–1140. [[CrossRef](#)]
13. Henriques, A.C.; Winkler, I. The advancement of virtual reality in automotive market research: Challenges and opportunities. *Appl. Sci.* **2021**, *11*, 11610. [[CrossRef](#)]
14. Checa, D.; Saucedo-Dorantes, J.J.; Osornio-Rios, R.A.; Antonino-Daviu, J.A.; Bustillo, A. Virtual reality training application for the condition-based maintenance of induction motors. *Appl. Sci.* **2022**, *12*, 414. [[CrossRef](#)]

15. Cordeiro, A.; Leite, R.; Gregory, L.; Catapan, M.; Siqueira, A.; Silva, T.; Brandão, M.; Ambrosio, P.; Reis, P.; Winkler, I. Immersive Technologies to Height Safety: Training Evaluation and Insights. In Proceedings of the 2025 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR), Lisbon, Portugal, 27–29 January 2025; pp. 435–440. [CrossRef]
16. Creswell, J.W.; Creswell, J.D. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*; SAGE Publications: Thousand Oaks, CA, USA, 2017.
17. Gregor, S.; Hevner, A.R. Positioning and presenting design science research for maximum impact. *MIS Q.* **2013**, *37*, 337–355. [CrossRef]
18. Offermann, P.; Blom, S.; Schönherr, M.; Bub, U. Artifact types in information systems design science—A literature review. In *Global Perspectives on Design Science Research: DESRIST 2010, St. Gallen, Switzerland*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 77–92.
19. Cordeiro, A.M.; Santos, F.; Winkler, I. Effectiveness of industrial training using virtual reality to mitigate risks associated with the work environment: A literature review. *Blucher Eng. Proc.* **2023**, *10*, 564–571. [CrossRef]
20. Slater, M.; Cabriera, C.; Senel, G.; Banakou, D.; Beacco, A.; Oliva, R.; Gallego, J. The sentiment of a virtual rock concert. *Virtual Real.* **2023**, *27*, 651–675. [CrossRef]
21. Samsung. Galaxy Watch7—Health Monitoring Specifications. 2024. Available online: <https://www.samsung.com/us/watches/galaxy-watch7/> (accessed on 27 August 2025).
22. Pico, X.R. Pico 4 Enterprise Specifications. 2022. Available online: <https://www.picoxr.com/global/products/pico4-enterprise> (accessed on 14 July 2025).
23. Venkatesh, V.; Davis, F.D. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Manag. Sci.* **2000**, *46*, 186–204. [CrossRef]
24. Venkatesh, V.; Morris, M.G.; Davis, G.B.; Davis, F.D. User acceptance of information technology: Toward a unified view. *MIS Q.* **2003**, *27*, 425–478. [CrossRef]
25. Witmer, B.G.; Jerome, C.; Singer, M.J. The factor structure of the presence questionnaire. *Presence* **2005**, *14*, 298–312. [CrossRef]
26. Silva, G.R.; Donat, J.C.; Rigoli, M.M.; de Oliveira, F.R.; Kristensen, C.H. A questionnaire for measuring presence in virtual environments: Factor analysis and Brazilian adaptation. *Virtual Real.* **2016**, *20*, 237–242. [CrossRef]
27. Grassini, S.; Laumann, K. Questionnaire Measures and Physiological Correlates of Presence: A Systematic Review. *Front. Psychol.* **2020**, *11*, 349. [CrossRef]
28. Archer, N.S.; Bluff, A.; Eddy, A.; Nikhil, C.K.; Hazell, N.; Frank, D.; Johnston, A. Odour enhances the sense of presence in a virtual reality environment. *PLoS ONE* **2022**, *17*, e0265039. [CrossRef]
29. Tavares, A.S.; Soares, M.M.; Marçal, M.A. Design and emotional responses: Is there coherence between what is said and what is felt? A study using biofeedback and virtual reality. *Virtual Real.* **2024**, *28*, 87. [CrossRef]
30. Meehan, M.; Razzaque, S.; Insko, B.; Whitton, M.; Brooks, F.P., Jr. Physiological reaction as an objective measure of presence in virtual environments. *Presence* **2002**, *11*, 538–549.
31. Spanlang, B.; Fröhlich, T.; Descalzo, F.F.; Antley, A.; Slater, M. The Making of a Presence Experiment: Responses to Virtual Fire. In Proceedings of the 10th Annual International Workshop on Presence, Barcelona, Spain, 25–27 October 2007.
32. Narciso, D.G. Realistic Multisensory Virtual Reality Simulator for Firefighter Training: Comparing the Physiological Response of Trainees in Real and Virtual Environments. Ph.D. Thesis, Universidade De Trás-Os-Montes e Alto Douro, Vila Real, Portugal, 2022.
33. Martens, M.A.; Antley, A.; Freeman, D.; Slater, M.; Harrison, P.J.; Tunbridge, E.M. It feels real: Physiological responses to a stressful virtual reality environment and its impact on working memory. *J. Psychopharmacol.* **2019**, *33*, 1264–1273. [CrossRef]
34. Slater, M.; Banakou, D.; Beacco, A.; Gallego, J.; Macia-Varela, F.; Oliva, R. A Separate Reality: An Update on Place Illusion and Plausibility in Virtual Reality. *Front. Virtual Real.* **2022**, *3*, 914392. [CrossRef]
35. Llinares, C.; Higuera-Trujillo, J.L.; Montañana, A. A Comparative Study of Real and Virtual Environment via Psychological and Physiological Responses. *Appl. Sci.* **2024**, *14*, 232. [CrossRef]
36. Haj-Bolouri, A.; Katende, J. Gamified immersive safety training in virtual reality: A case from the railway industry. *J. Workplace Learn.* **2024**, *36*, 271–288. [CrossRef]
37. Lee, J.; Kim, D.; Park, Y. Effects of point-and-level based gamification on motivation and performance in a VR assembly task. *J. Comput. Assist. Learn.* **2022**, *38*, 1205–1216. [CrossRef]
38. Carnell, S.; De Siqueira, A.G.; Miles, A.; Lok, B. Informing and evaluating educational applications with the Kirkpatrick model in virtual environments: Using a virtual human scenario to measure communication skills behavior change. *Front. Virtual Real.* **2022**, *3*, 810797. [CrossRef]
39. Winkler, I.; Ambrósio, P.E.; Leite, R.M.C.; Cordeiro, A.M.; Almeida, L.G.G.; Thasla, Y.; Siqueira, A.G.; Catapan, M.F.; Berretta, L.O. Biofeedback na avaliação da experiência do usuário em ambientes imersivos. In Proceedings of the Anais do 25º Simpósio Brasileiro de Computação Aplicada à Saúde—SBCAS 2025, Porto Alegre, Brazil, 9–13 June 2025.
40. Makransky, G.; Terkildsen, T.S.; Mayer, R.E. Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learn. Instr.* **2019**, *60*, 225–236. [CrossRef]

41. Nasri, M.; Kosa, M.; Chukoskie, L.; Moghaddam, M.; Harteveld, C. Exploring eye tracking to detect cognitive load in complex virtual reality training. In Proceedings of the 2024 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Bellevue, WA, USA, 21–25 October 2024; IEEE: New York, NY, USA, 2024; pp. 51–54. [[CrossRef](#)]
42. Doolani, S.; Wessels, C.; Kanal, V.; Sevastopoulos, C.; Jaiswal, A.; Nambiappan, H.; Makedon, F. A review of extended reality (XR) technologies for manufacturing training. *Technologies* **2020**, *8*, 77. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.