Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Designing Feedback-Supported Immersive Virtual Reality for Safety Training

Idowu Awoyemi

idawoyemi@crimson.ua.edu

The University of Alabama

Jewoong Moon

The University of Alabama

Siyuan Song

Arizona State University

Research Article

Keywords: construction safety, experiential learning, formative evaluation, instructional design, immersive virtual reality

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11724809/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Designing Feedback-Supported Immersive Virtual Reality for Safety Training

Abstract

Immersive virtual reality (IVR) lets learners rehearse safety decisions, but realistic scenes alone do not make effective instruction. This article reports the design and formative evaluation of an IVR construction safety environment structured around experiential learning theory. A pilot with five students led to revisions in scenario boundaries, task framing, feedback and session length. A mixed-methods evaluation with 30 undergraduate construction engineering students then combined an immersion survey, observer checklists and think-aloud data. Participants reported very high attention and moderate emotional involvement, most showed high controller proficiency, and qualitative accounts suggested that learners valued explanations linking hazards to safety principles. Controller onboarding and facilitation remained implementation demands. We translate the development record into four recommendations for practitioners: constrain complexity, connect actions with explanatory feedback, clarify goals and controls, and plan comfort and facilitation. Learning gains and transfer were not measured.

Keywords: construction safety; experiential learning; formative evaluation; instructional design; immersive virtual reality

Introduction

Hazard identification is a central task in construction safety training. Immersive virtual reality (IVR) provides a simulated setting in which learners can encounter and assess construction hazards. Sacks et al. (2013) evaluated this approach against classroom training with visual aids, finding benefits for some construction activities but not for general site safety. For instructional designers, an important design question is how to connect simulated hazard encounters with clear goals, explanatory feedback, and opportunities to act on that feedback.

This study reports the design and formative evaluation of an IVR safety training environment for undergraduate construction engineering students. Kolb's Experiential Learning Theory guided the design of feedback linking hazard encounters with reflection, safety explanations and revised actions (Kolb, 1984; Kolb & Kolb, 2018). A pilot with five students informed changes to the prototype; a subsequent evaluation with 30 students examined learner experience and implementation constraints. The resulting recommendations concern how to design and facilitate immersive learning activities, with construction safety providing the application context.

Two questions guided the evaluation: (1) How do students perceive the usability, engagement and instructional value of the IVR training environment? (2) What implementation challenges and operational constraints emerge during deployment? We use the findings to explain design decisions and propose practical recommendations for educational technology practitioners. The evaluation establishes formative evidence about the prototype; it does not test instructional effectiveness, learning gains or transfer to workplace performance.

Design Rationale

Connecting Immersion with Guided Practice

IVR permits learners to encounter hazards and rehearse responses without exposure to real construction risks (Sacks et al., 2013). Reviews of educational VR identify promising applications while emphasizing the importance of instructional design and clearer evidence about learning processes (Merchant et al., 2014; Radianti et al., 2020). Realistic environments therefore require instructional decisions about what learners should notice, how they should respond and what support they receive. Presence and engagement can help sustain participation, but they should not be treated as evidence of learning. Designers must consider whether navigation, visual detail and interaction demands compete with the task itself (Makransky & Petersen, 2021). For novice users, a bounded practice environment may provide a more usable introduction than unrestricted exploration of a complex virtual site.

Feedback is central to this design problem. Confirmation tells learners whether an action registered; explanation connects that action to a safety rationale. Formative feedback should give learners information they can use to reconsider and improve their responses (Shute, 2008). In an immersive environment, the timing and presentation of this information must also accommodate controller use, visual attention and physical comfort. Our design therefore prioritized documenting the relationship among learning goals, feedback features and observed use. This account makes the rationale for prototype revisions visible and identifies decisions practitioners can consider when adapting immersive activities to their own contexts.

For this development project, the design problem was how to connect immersive hazard encounters with structured opportunities for reflection. Research on virtual safety-training design and experiential learning provides relevant context for examining how these experiences are

organized (Speiser & Teizer, 2024; Tonteri et al., 2023). Two needs motivated the present implementation: integrating feedback that connects hazard encounters with safety explanations and subsequent action, and documenting how formative evidence informs that support.

Using Experiential Learning to Structure Feedback

Kolb's learning cycle provided a design framework: concrete experience involved encountering a hazard; reflective observation involved pausing to consider the response; abstract conceptualization involved connecting the hazard to a safety explanation; and active experimentation involved applying or retrying an action (Kolb, 1984; Kolb & Kolb, 2018). These stages describe intended instructional opportunities rather than measured cognitive transitions. Figure 1 presents our synthesis of experiential learning and situated cognition as a framework for the prototype features. In the prototype, audio-visual confirmation cues, explanatory overlays, a post-feedback pause and facilitator questions provided opportunities to consider the hazard and response, while OSHA-related explanations linked particular situations to broader safety principles. Mitigation actions and scenario retries provided opportunities to use that information. The evaluation examined participants' experience of this sequence rather than validating the theory.

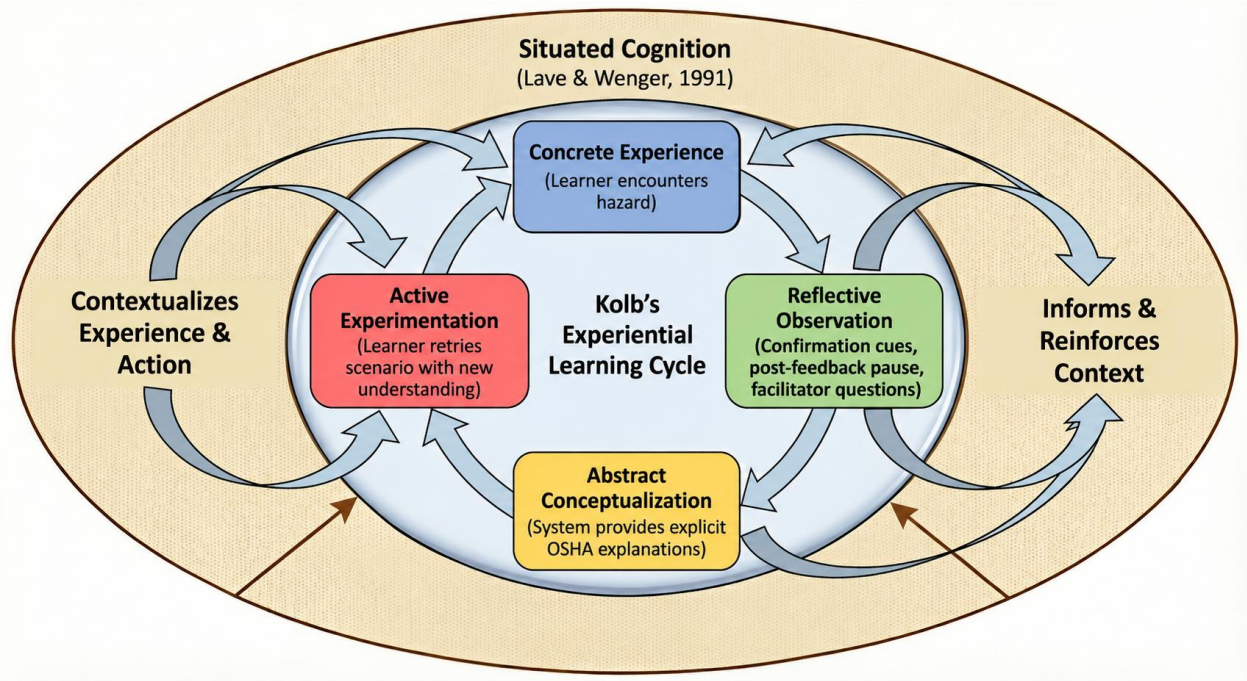


Figure 1. Authors' synthesis of experiential learning, situated cognition and prototype features (Kolb, 1984; Kolb & Kolb, 2018; Lave & Wenger, 1991).

Situated learning further informed the use of recognizable work settings and safety practices (Lave & Wenger, 1991). Authenticity was treated selectively: the environment needed to preserve the meaning of the safety task while remaining manageable for novice headset users. Together, these perspectives guided decisions about scenario boundaries, explanatory feedback and facilitator support.

Design and Formative Development

We used design and development research (DDR) to document the creation, refinement and formative evaluation of an instructional product (Richey & Klein, 2007). The four phases were needs analysis and context examination, initial design, formative pilot testing and refinement, and final design specifications (Figure 2). The practical contribution is a traceable

account of how learner feedback and observation informed design decisions. Recommendations derived from this local development process remain provisional and require evaluation in other settings.

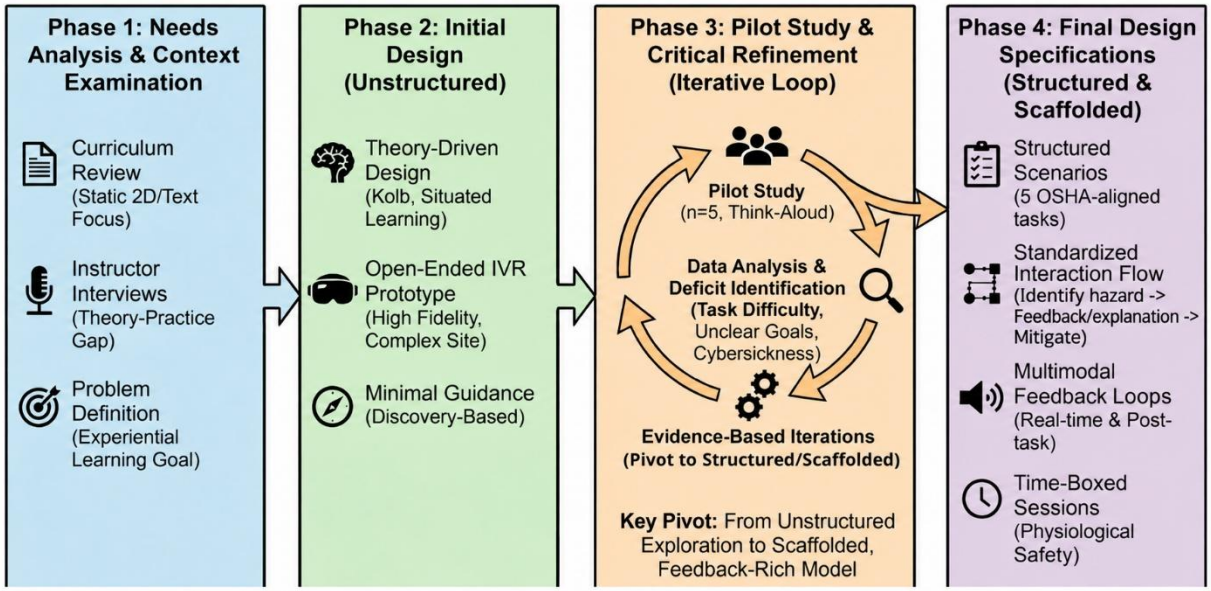


Figure 2. Design and development cycle across the four DDR phases, with hazard identification followed by feedback/explanation and then mitigation in the final prototype.

Phase 1: Needs Analysis and Context Examination

Phase 1 focused on clarifying the instructional problem and local constraints. A review of the undergraduate construction engineering safety curriculum revealed heavy reliance on text-based OSHA regulations and static 2D images for hazard identification training. Consultations with Construction Safety Management course instructors confirmed that students could often recite safety regulations but struggled to apply them in complex, dynamic construction scenarios. This analysis led to the primary design goal: to create an IVR environment that bridges theoretical safety knowledge and real-time hazard identification by situating learning in simulated construction scenes and embedding structured opportunities for reflection.

Phase 2: Initial Design

Guided by the Phase 1 findings, the initial IVR prototype was developed as a high-fidelity, open-ended construction site. Kolb and Kolb's (2018) account of concrete experience informed the provision of simulated hazard encounters, while situated learning informed the use of recognizable work settings and safety practices (Lave & Wenger, 1991). The design team initially prioritized authenticity and unrestricted exploration: learners could freely navigate a multi-area virtual site containing numerous potential hazards, without explicit checklists or detailed guidance. This approach was intended to encourage exploration and reveal difficulties in hazard recognition; its suitability was examined through the subsequent formative pilot. The prototype featured scaffolding, electrical, and material-storage work zones. Learners were instructed to "inspect the site and identify safety hazards," approximating a construction-site walkthrough.

Phase 3: Pilot Study and Critical Refinement

Pilot testing identified difficulties with unstructured exploration, unclear goals, limited feedback and physical comfort. We responded by bounding the search areas, clarifying tasks, introducing multimodal feedback and limiting scenario duration.

To examine these learner-level responses, we conducted a formative pilot study with a convenience sample of five construction engineering students who were not part of the main study. The pilot followed a rapid iterative testing protocol (Medlock et al., 2005). Each participant was briefed on the "open exploration" task and instructed to identify as many hazards as possible within the unstructured Phase 2 environment, with a maximum session cap of 20 minutes. Participants utilized a concurrent think-aloud protocol to verbalize their thought processes and frustrations in real-time. Immediately following the session, we conducted brief, semi-structured debriefing interviews focused on clarity, cognitive workload, physical comfort,

and perceived learning value. We analyzed field notes, observation checklists, and interview transcripts using a rapid thematic analysis approach to identify critical friction points and pedagogical breakdowns, which directly informed the transition to the structured design.

The pilot analysis identified five recurring problems: unfocused scanning, difficulty managing the number of potential hazards, uncertainty about task completion, limited perceived instructional value without feedback, and discomfort during prolonged exploration. We treated these as design problems requiring scaffolding, and they informed the revisions summarized in Table 1.

Table 1 Evidence-Based Iterations during Phase 3

Phase 2 Design Decision	Learner difficulty observed in the pilot	Targeted Design Modification	Practical Recommendation
Unstructured exploration in an open-world environment	Unfocused scanning; fixation on a single hazard category	Structured scenarios with bounded search areas and predefined hazard zones	Recommendation 1: Constrain Complexity
Large, undefined number of potential hazards	Difficulty managing task demands and incomplete hazard coverage	Reduced hazard complexity and structured scenarios	Recommendation 1: Constrain Complexity
Implicit task framing (“explore and find hazards”)	Uncertainty about task completion and aimless navigation	Explicit task definition with clear goals.	Recommendation 3: Clarify Goals and Controls
Delayed or absent feedback	Uncertainty about performance and limited perceived instructional value	Multimodal feedback loops, including real-time cues and post-task explanations	Recommendation 2: Connect Actions With Explanatory Feedback
Prolonged exploration within the pilot session cap	Physical discomfort and disengagement during prolonged exploration	Time-boxed scenarios with capped headset duration	Recommendation 4: Plan Comfort and Facilitation

Phase 4: Final Design Specifications

Phase 4 consolidated the iterative changes into the final IVR training environment. Figure 3 contrasts the initial open layout with the structured approach developed in Meta's Horizon Worlds. The five scenarios addressed (a) slip and fall hazards, (b) unsafe scaffolding placement and falls from elevation, (c) struck-by hazards involving falling objects, (d) caught-in/between hazards involving heavy equipment, and (e) electrical hazards. The scenario set drew on OSHA's Construction Focus Four categories: falls, struck-by, caught-in/between, and electrocution (Occupational Safety and Health Administration [OSHA], n.d.). Slip and fall hazards and scaffold-related falls were presented as separate fall-related situations; the five scenarios therefore did not represent five distinct Focus Four categories.

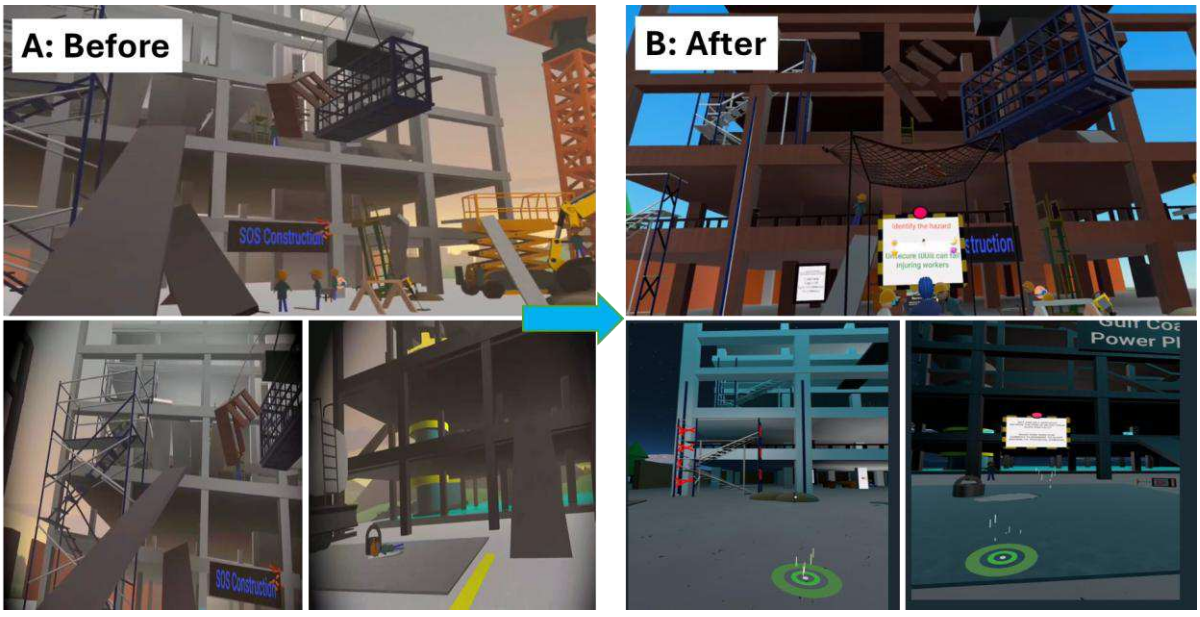


Figure 3. Evolution from the initial open exploration layout to structured scenarios with bounded search areas and explicit task cues

The revised interaction sequence linked navigation and identification with feedback, explanation and mitigation. Learners entered bounded scenarios to locate hazards. Identification or time expiry triggered audio-visual feedback, followed by video overlays or text explaining the relevant safety guidance. Learners then applied an appropriate response, such as placing a caution sign near a slippery surface. This sequence was designed to give learners repeated opportunities for reflection and application.

Evaluation Study Methods

We conducted a convergent mixed-methods formative evaluation, collecting survey responses, observational records and think-aloud data within individual sessions (Creswell & Plano Clark, 2018). Descriptive statistics characterized self-reported immersion and recorded controller use. Qualitative analysis examined perceived instructional value and deployment constraints; the strands were integrated during interpretation.

Recommendation Development and Data Mapping

We mapped pilot observations and main-evaluation findings to the design rationale to develop provisional recommendations. Table 2 connects the evidence with theoretical interpretations and practical implications. The pilot informed prototype changes, whereas the main evaluation identified remaining strengths and constraints. Because the samples and designs differed, these phases were not treated as a controlled before-and-after comparison.

Table 2

Mapping Observations to Interpretations and Recommendations

Data Source & Key Finding	Theoretical Interpretation	Practical Recommendation
Pilot observations indicated unfocused scanning, difficulty managing task demands and discomfort.	Bounded tasks were intended to make the initial experience more manageable.	Recommendation 1: Constrain Complexity
Pilot participants reported limited instructional value without explanatory feedback; main-evaluation accounts valued the explanations.	Feedback was intended to create opportunities for reflection and connecting actions to safety rationales.	Recommendation 2: Connect Actions With Explanatory Feedback
Observer notes recorded difficulty using controls and manipulating objects; pilot accounts identified uncertainty about task goals.	Clear goals and controls were intended to support purposeful action.	Recommendation 3: Clarify Goals and Controls
Technical assistance was needed in some main-evaluation sessions; pilot accounts identified dizziness.	Physical conditions and technical support affected access to the activity; effects on learning were not measured.	Recommendation 4: Plan Comfort and Facilitation

Participants

Following IRB approval (Protocol ID: 24-01-7255), we recruited 30 participants (males = 18; females = 12) via convenience sampling from undergraduate construction engineering students at a large public university. Inclusion criteria required participants to be (a) at least 18 years old and (b) currently enrolled in or having recently completed the Construction Safety Management course, aligning recruitment with the course context examined in the needs analysis. All 30 participants completed the protocol (0% attrition). Each participant attended a

50–60-minute individual lab session following the study protocol. First, researchers explained the study, obtained informed consent, and collected demographics (age, academic year, prior IVR experience). Participants were then introduced to the Meta Quest 2 and controllers and completed a think-aloud practice exercise (e.g., counting windows in their home) to acclimate to the protocol. Following an in-headset tutorial, participants began the main IVR training task (approximately 25–30 minutes) using think-aloud throughout. Upon completion, participants removed the headset and completed the Immersion in IVR Survey via Qualtrics. Sessions concluded with brief, unstructured retrospective discussions (5–10 minutes) probing observations, followed by debriefing.

Instrumentation

Participants completed a post-session Immersion in IVR Survey informed by Cheng and Tsai (2020). The local adaptation used four dimensions (Basic Attention, Temporal Dissociation, Transportation and Emotional Involvement), with four items per dimension and a five-point agreement scale (1 = strongly disagree, 5 = strongly agree). Appendix A reproduces the 16 items and maps them to the exported variables. The source study also included enjoyment, which was not included in this local instrument. These scores describe self-reported immersion, not learning.

The researcher-completed IVR Usability Testing Checklist documented preparation, controller use and navigation, hazard-related activity, user experience and feedback (Appendix B). This report retains the controller-proficiency and navigation-difficulty categories recorded in 27 checklist records. Behaviors recorded under the checklist's "cognitive overload" label, such as frustration or requests for help, were interpreted as difficulty managing task demands. Appendix C defines the retained variables.

Concurrent think-aloud protocols, headset recordings and observer notes contextualized survey responses. Identification meant locating or selecting a hazard; mitigation meant carrying out a subsequent response, such as placing a caution sign. Identification and mitigation were not scored as performance outcomes in this evaluation.

Data Analysis

Quantitative Analysis

We summarized the Immersion in IVR Survey responses for 30 participants. Dimension scores were the mean of available items for each participant, followed by the cohort mean and standard deviation. Two participants had three rather than four Basic Attention responses; a complete-case check produced nearly identical results ($M = 4.75$, $SD = 0.77$, $n = 28$). Controller proficiency and navigation difficulty were summarized as category counts and percentages from 27 checklist records. Hazard accuracy and task-time data were not analyzed because the scoring protocol was not standardized across sessions; this evaluation therefore focuses on learner experience and implementation.

Qualitative Analysis

Qualitative analysis focused on think-aloud session data. Fifteen of the 30 sessions were transcribed verbatim. The team coded transcripts in sequence and monitored thematic saturation: no new codes or themes emerged after the ninth transcript, and the next three served as confirmatory cases, so formal analysis comprised 12 transcripts (Hennink & Kaiser, 2022). These transcripts are labelled P1 to P12 in transcript order; the labels are not checklist IDs. The same session recordings also informed a separate analysis of learning processes.

We used inductive thematic analysis to examine usability, feedback value and engagement (Braun & Clarke, 2006). The research team read and coded transcripts, compared interpretations, refined the codebook and discussed discrepancies, consulting video records and

observer notes for context. To support trustworthiness, a second researcher independently coded 25% (n = 3) of the transcripts, and negotiated agreement, calculated as the percentage of coded segments assigned the same code after the coders discussed discrepancies, exceeded 85%.

Mixed-Methods Integration

We compared the survey summaries with qualitative accounts and observational notes during interpretation. Self-reported attention and immersion were considered alongside controller difficulties, requests for support and comments about explanations. This integration identified implementation issues and perceived value without treating favorable immersion ratings as evidence of learning.

Results

Findings address learner experience and implementation constraints. Survey responses describe immersive engagement, categorical checklist records describe controller use and navigation difficulty, and qualitative accounts identify perceived value and remaining design problems.

Participant Characteristics

The study sample comprised 30 construction engineering undergraduate students currently enrolled in or having recently completed "Construction Safety Management," providing prior exposure to construction safety concepts. The cohort included sophomores, juniors, and seniors, aged 19 to 24 years (M = 21.2, SD = 1.4), with 60% male and 40% female participants. Prior IVR experience varied: no experience (n = 12, 40%), limited experience (n = 15, 50%), or frequent usage (n = 3, 10%). The study retained 100% of participants (0% attrition) throughout the protocol; however, technical recording errors corrupted video files for three participants. The retained checklist summaries cover 27 participants; self-report measures include all 30. Figure 4 illustrates participants engaging with the IVR system during the data collection sessions.



Figure 4. Participants engaging in the IVR safety training session

RQ1 Findings: Perceptions of Usability, Engagement, and Instructional Value

We examined survey responses, observational records and qualitative themes to characterize usability, engagement and perceived instructional value.

Engagement

Basic Attention received the highest mean rating ($M = 4.77$, $SD = 0.75$) and Emotional Involvement the lowest ($M = 3.27$, $SD = 0.96$). Temporal Dissociation and Transportation fell between them (Table 3).

Table 3

Descriptive statistics for immersive engagement

Dimension	n	M	SD
Basic Attention	30	4.77	0.75
Temporal Dissociation	30	3.40	0.83
Transportation	30	3.56	1.01

Dimension	n	M	SD
Emotional Involvement	30	3.27	0.96

Usability

Of the 27 checklist records, 23 (85.2%) rated controller proficiency as High or Excellent. Navigation difficulty was recorded as None or Minor for 24 participants (88.9%); Moderate, Significant and Major difficulty were each recorded once (Table 4).

Observed Interaction with the Prototype

Observer notes illustrate the interface demands behind these categories, for example "Forgot side button initially; adapted well" and "Struggled with object manipulation; required guidance." These notes indicate where orientation and technical assistance were needed.

Table 4

Controller proficiency and navigation-difficulty categories

Measure	Recorded category	n	%
Controller proficiency	Low	1	3.7
Controller proficiency	Moderate	3	11.1
Controller proficiency	High	20	74.1
Controller proficiency	Excellent	3	11.1
Navigation difficulty	None	14	51.9
Navigation difficulty	Minor	10	37.0
Navigation difficulty	Moderate	1	3.7
Navigation difficulty	Significant	1	3.7
Navigation difficulty	Major	1	3.7

Note. N = 27 checklist records. Percentages use 27 as the denominator within each measure and may not total 100 because of rounding. Category names reproduce the archived checklist; "None" is an observed category. No numerical spacing between severity categories is assumed.

Qualitative Perceptions of Instructional Value

We report themes from the 12 coded transcripts. Much of the verbalization responded to facilitator questions, so counts indicate how many participants expressed a pattern at least once, not how often it occurred spontaneously. Three themes emerged.

Articulating a safety rationale. When asked why a step or response was needed, 10 of 12 participants gave a reason rather than repeating the instruction. Explaining the lockout/tagout sequence, P9 observed that "Placing a lock on it already makes a decision for people that they can't use it," and P10 reasoned that "turning off the power eliminates the safety issue at its root." P5 explained the final step: "even if it's partially removed, there's still the electrical current jumping between the plug and the outlet." The two participants who did not articulate a rationale read the steps aloud as a procedure (e.g., "Shut off power. Sign out lock," P1). A main-evaluation observer note likewise recorded that one participant "articulated safety rationale clearly."

Evaluating and extending the feedback. Seven participants explicitly judged the feedback, usually when the facilitator asked whether they agreed with it. P8 described it as accurate and acknowledged a missed cue: "I mean, it was accurate. I didn't see it up there." P12, asked whether they agreed with the mitigation feedback, replied: "Strongly agree with that." P10 extended it: "I think it was a great feedback. Um, the only thing I might add is, like, making sure that everyone is aware of what's going on at the site, not just relying on, like, those safety nets."

Relating the scenarios to work and prior training. Two participants connected the scenarios to prior training or future site work. P5, asked about the value of the training, contrasted an internship safety course that "was all text and reading and whatnot" and suggested that "virtual reality is a great way to learn and actually see the situation played out." P8, who had not visited a construction site, noted: "People like me have never been to site."

Expressions of enjoyment appeared in at least four sessions (e.g., "Uh, that's cool," P6), consistent with the high Basic Attention ratings.

RQ2 Findings: Implementation Challenges and Deployment Constraints

Thematic analysis of the think-aloud transcripts, triangulated with observer notes, identified three categories of implementation demands: (a) technical and hardware issues, (b) physical and human factors and (c) instructional and logistical constraints.

Technical and Hardware Challenges

Controller use and technical interruptions affected some sessions. Six of 12 think-aloud participants asked how to perform a controller action or needed step-by-step guidance to grab, select or place objects (e.g., "I'm holding it, and then what do I click?" P9; "How do I place?" P2). One scenario stopped responding and had to be reset ("Not working," P2), low feedback volume was raised in three sessions (e.g., "Should I turn the volume up," P6). Checklist notes recorded similar difficulties with object manipulation and controller buttons. Because the facilitator also supported the think-aloud procedure, technical assistance and instructional prompting were not logged separately.

A related interaction issue concerned the enforced pause after feedback. To give learners time with each explanation, the system held them in place for several seconds after a correct identification. Four participants remarked on being unable to move (e.g., "how do I escape it?" P7; "it's not moving around," P8), and the facilitator explained the pause in several other

sessions. The pause was intended to support reflection, but without a visible cue some learners experienced it as a malfunction.

Physical and Human Factors

Pilot participants reported dizziness during extended, unstructured exploration, which led to bounded scenarios and capped headset time. In the main evaluation, no physical discomfort was recorded in any of the 27 checklist records, although this record relied on observable signs and participant report.

Instructional and Logistical Implementation

This implementation used individual 50–60-minute sessions encompassing orientation, the approximately 25–30-minute training activity, survey completion and debriefing. The facilitator administered the think-aloud protocol, monitored the physical setting and assisted with technical difficulties. The facilitator also asked follow-up questions after most scenarios (e.g., why a step was necessary), and these questions elicited much of the safety reasoning reported above. Classroom scheduling, shared-device use and more independent delivery would require separate evaluation.

Discussion

Participants reported very high attention (M = 4.77) and more moderate emotional involvement and transportation (M = 3.27 to 3.56). These descriptive ratings characterize participants' experience of this implementation; they do not establish a loss of presence attributable to bounded scenarios or explicit tasks. Presence and learning are distinct constructs (Makransky & Petersen, 2021), and learning gains were not measured here. In the think-aloud analysis, 10 of 12 participants articulated a safety rationale when prompted. This observation documents reasoning expressed during facilitated use, but does not establish that participants

engaged more with explanations than with other features or that feedback produced those rationales.

A second lesson concerns where the design problems came from. In the pilot, the most consequential difficulties were about orientation rather than safety content: learners did not know where to look, how many hazards to expect or when they were done. This matches guidance that novices benefit from reduced complexity and explicit task structure before full-task practice (van Merriënboer & Kirschner, 2017). IVR safety training therefore inherits a familiar instructional design problem, and the bounded scenarios, explicit goals and explanatory feedback were conventional instructional moves applied to an unconventional medium.

Third, deployment conditions belong in the design account. Controller onboarding, facilitator support and session length shaped how learners could engage with the activity. Two observations stand out. The enforced pause after feedback, intended to create room for reflective observation, was experienced by some participants as being stuck; a design feature serving one stage of Kolb's cycle can fail if learners cannot tell why it is happening. And much of the safety reasoning in the transcripts emerged in response to facilitator questions, which suggests that the questioning, not only the system feedback, carried part of the reflective work. Sacks et al. (2013) found that the advantages of IVR over classroom training varied across construction activities; our observations suggest that the practical conditions of headset use are one place to look for such variation. The recommendations below combine the documented iterations with these remaining constraints. Because several changes were introduced together, they are best read as a set.

Practical Design Recommendations

The recommendations translate established instructional principles into decisions about immersive interfaces, feedback and facilitation. Each draws on evidence from this development process and on prior literature, and each should be tested in local contexts.

Recommendation 1: Constrain Complexity

Preserve task-relevant authenticity while limiting unnecessary complexity. In the pilot, the unrestricted environment produced unfocused scanning, fixation on a single hazard category and incomplete coverage. The revised prototype used bounded scenarios and defined hazard zones (Table 1), and in the main evaluation 24 of 27 participants experienced no more than minor navigation difficulty.

Selective realism and the gradual introduction of task complexity provide useful design precedents (Alessi & Trollip, 2001; van Merriënboer & Kirschner, 2017). A designer should retain the environmental details needed to interpret a hazard while examining whether additional objects or interactions distract from the intended task.

Begin local testing with a manageable set of hazards and clearly bounded search areas. Observe where learners hesitate, become disoriented or attend to irrelevant objects. Increase complexity only when the task remains clear.

Recommendation 2: Connect Actions With Explanatory Feedback

Pair action feedback with an explanation learners can use. Pilot participants found limited value in exploration without feedback. The revised design combined confirmation cues with explanations of the safety rationale and an opportunity to carry out a mitigation action, and in the main evaluation seven of 12 think-aloud participants explicitly agreed with or extended the feedback (Table 2). Kolb's cycle helped us locate feedback between the initial encounter and a subsequent action (Kolb, 1984). Designers can use it to ask what learners need to notice, what explanation they need and what they can do next.

In practice, distinguish confirmation from explanation. A highlight or sound can acknowledge a selection; concise narration, text or video can explain the risk and an appropriate response. Keep these elements coordinated so learners can attend to the explanation without simultaneously managing complex movement. Provide a chance to apply the feedback and check that the next action is understandable (Shute, 2008).

Recommendation 3: Clarify Goals and Controls

Make the objective, controls and completion criteria explicit. Pilot participants were uncertain about what counted as completing the inspection. The revised prototype introduced clearer framing and feedback, but some uncertainty remained in the main evaluation (e.g., “did I select the wrong one?” P1; “is it still this one or do I go?” P3), and six of 12 think-aloud participants needed controller guidance, indicating a continuing need for orientation in this implementation. A briefing should distinguish the safety task from the interface task: learners need to know both what counts as a hazard and how to register a response. Pre-training and signaling provide the instructional basis for this separation (Mayer, 2014). Explain the task without disclosing the answers the activity is meant to elicit.

Use an interactive controller tutorial, concise scenario instructions and visible system-state cues. During formative testing, ask learners to explain the goal and show how they will select an object before the main activity begins. Record whether requests for help concern controls, task rules or safety reasoning; these problems require different design responses. Progress indicators and accessible help are possible extensions, but their form should be tested with the intended learners. Signal enforced waits explicitly; in this study an unexplained post-feedback pause was read as a malfunction. Avoid assuming that a persistent counter or additional overlay will improve performance: extra interface information can also compete for attention

(Makransky & Petersen, 2021; Mayer, 2014). The purpose is to reduce procedural uncertainty while preserving meaningful hazard decisions.

Recommendation 4: Plan Comfort and Facilitation

Plan orientation, technical assistance and pause procedures alongside the virtual activity. Pilot reports of dizziness led to time-boxed scenarios, and no discomfort was recorded in the main evaluation. Each session still required a facilitator to orient the participant, monitor the physical space and resolve interaction problems. Facilitator questions also elicited much of the safety reasoning observed; if the activity is to run without a facilitator, comparable reflection prompts should be built into the system. The session durations reported here describe this protocol, not recommended exposure limits.

Before classroom adoption, plan equipment checks, clear physical space, accessible stopping procedures, orientation time and technical assistance. A shared-device activity may require rotations or a different scheduling model. Learners who cannot comfortably use a headset should have an alternative route to the instructional objective. These are implementation recommendations arising from the observed constraints, rather than evaluated delivery models.

Implications for Practitioners

The four recommendations can structure a local design review. Table 5 shows how each stage of the intended learning cycle was implemented in the prototype and which recommendation supports it. For another immersive activity, the transferable question is whether each design feature addresses a documented learner need and gives learners an opportunity to act on feedback.

Table 5

Kolb stages, prototype features and supporting recommendations

Kolb Stage	Situated Cognition	Practical Recommendation	IVR Implementation
Concrete Experience	Authentic practice context	Recommendation 1: Constrain Complexity	Bounded scenarios with focused hazard zones
Reflective Observation	Expert guidance	Recommendation 2: Connect Actions With Explanatory Feedback	Audio and visual confirmation cues, a post-feedback pause and facilitator questions about the hazard and safety response
Abstract Conceptualization	Cultural tools and norms (OSHA guidance)	Recommendation 2: Connect Actions With Explanatory Feedback	Video and text explanations linking the hazard to OSHA guidance
Active Experimentation	Facilitated participation	Recommendation 3: Clarify Goals and Controls	Mitigation actions (e.g., placing a caution sign) and scenario retries
All stages	Facilitated participation	Recommendation 4: Plan Comfort and Facilitation	Orientation, time-boxed scenarios and facilitator support

A practical review sequence follows from this account. First, specify the decision learners must make and the visual information needed for it. Second, test whether learners can navigate and register a response before interpreting any task scores. Third, check whether learners can describe an appropriate next action in their own words after the explanation. Fourth, log the support and time each session requires, recording technical assistance separately from instructional prompting. Finally, document each revision as a trail from observed problem to

change to subsequent evidence, as Table 1 does here (McKenney & Reeves, 2019; Richey & Klein, 2007). This makes it easier to tell whether a difficulty concerns the interface, the task framing, the explanation or the delivery conditions.

Limitations and Future Research

The convenience sample of 30 students from one university limits transferability. Survey analyses were descriptive and used an adapted immersion instrument. Qualitative analysis covered 12 of 15 transcribed sessions, coded to saturation, and categorical checklist summaries covered 27 records. Much of the think-aloud reasoning was elicited by facilitator questions rather than produced spontaneously. The retained evidence supports a formative design account, not validation of the instrument or recommendations.

Because the hazard scoring protocol was not standardized, task accuracy and speed were not reported, and categorical navigation entries did not support event-level analysis. Observed difficulty managing task demands was not a direct measure of cognitive load, and technical assistance was not logged separately from instructional prompting, which limits conclusions about independent performance.

No pretest, comparison group, delayed assessment or workplace transfer measure was included. The pilot and main evaluation involved different samples, and multiple features changed together. Neither improvement between versions nor causal effects of individual features can therefore be established. Future studies should distinguish identification from mitigation, record first attempts separately from feedback and retries, define timing boundaries, and document rater sampling and agreement before interpreting performance.

Further development should prioritize reliable recordings, clearer documentation of facilitator interventions and accessibility. Subsequent studies could compare feedback sequences, examine changes in scenario complexity and assess whether learners apply explanations in new

tasks. Classroom trials should also document scheduling, equipment sharing and the resources required for routine use.

Conclusion

This design and formative evaluation offers practical guidance for developing feedback-supported IVR activities. Pilot observations informed bounded scenarios, clearer task framing and explanatory feedback; the main evaluation documented self-reported immersion, controller use and continuing implementation constraints. The contribution is a documented design process and four provisional recommendations for practitioners. Future evaluation should test learning, retention and transfer beyond this prototype.

References

Alessi, S. M., & Trollip, S. R. (2001). *Multimedia for learning: Methods and development* (3rd ed.). Allyn & Bacon.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Cheng, K.-H., & Tsai, C.-C. (2020). Students' motivational beliefs and strategies, perceived immersion and attitudes towards science learning with immersive virtual reality: A partial least squares analysis. *British Journal of Educational Technology, 51*(6), 2140–2159. https://doi.org/10.1111/bjet.12956

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.

Hennink, M., & Kaiser, B. N. (2022). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science & Medicine, 292*, Article 114523. https://doi.org/10.1016/j.socscimed.2021.114523

Kolb, A. Y., & Kolb, D. A. (2018). Eight important things to know about the experiential learning cycle. *Australian Educational Leader, 40*(3), 8–14. https://search.informit.org/doi/10.3316/informit.192540196827567

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice Hall.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press. https://doi.org/10.1017/CBO9780511815355

Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality.

Educational Psychology Review, 33(3), 937–958. https://doi.org/10.1007/s10648-020-09586-2

Mayer, R. E. (Ed.). (2014). *The Cambridge handbook of multimedia learning* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369

McKenney, S., & Reeves, T. C. (2019). *Conducting educational design research* (2nd ed.). Routledge. https://doi.org/10.4324/9781315105642

Medlock, M. C., Wixon, D., McGee, M., & Welsh, D. (2005). The rapid iterative test and evaluation method: Better products in less time. In R. G. Bias & D. J. Mayhew (Eds.), *Cost-justifying usability: An update for an Internet age* (2nd ed., pp. 489–517). Morgan Kaufmann. https://doi.org/10.1016/B978-012095811-5/50017-1

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education, 70*, 29–40. https://doi.org/10.1016/j.compedu.2013.07.033

Occupational Safety and Health Administration. (n.d.). *Construction focus four training*. U.S. Department of Labor. https://www.osha.gov/training/outreach/construction/focus-four

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education, 147*, Article 103778. https://doi.org/10.1016/j.compedu.2019.103778

Richey, R. C., & Klein, J. D. (2007). *Design and development research: Methods, strategies, and issues*. Routledge. https://doi.org/10.4324/9780203826034

Sacks, R., Perlman, A., & Barak, R. (2013). Construction safety training using immersive virtual reality. *Construction Management and Economics, 31*(9), 1005–1017. https://doi.org/10.1080/01446193.2013.828844

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Speiser, K., & Teizer, J. (2024). Formalizing virtual construction safety training: A schematic data framework enabling real-world hazard simulations using BIM and location tracking. *Journal of Information Technology in Construction, 29*, 980–1004. https://doi.org/10.36680/j.itcon.2024.043

Tonteri, T., Holopainen, J., Lumivalo, J., Tuunanen, T., Parvinen, P., & Laukkanen, T. (2023). Immersive virtual reality in experiential learning: A value co-creation and co-destruction approach. In T. X. Bui (Ed.), *Proceedings of the 56th Annual Hawaii International Conference on System Sciences* (pp. 1313–1322). Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2023.162

van Merriënboer, J. J. G., & Kirschner, P. A. (2017). *Ten steps to complex learning: A systematic approach to four-component instructional design* (3rd ed.). Routledge. https://doi.org/10.4324/9781315113210

Appendix A

Immersion in IVR Survey: Administered Item Wording

The item wording below reproduces the archived Qualtrics response export. BA, TD, TRA and EI codes are reporting labels; Q.1_1–Q.1_4, Q2_1–Q2_4, Q3_1–Q3_4 and Q4_1–Q4_4 are the corresponding exported variables. All four dimensions used the same five-point response scale: 1 = Strongly Disagree; 2 = Somewhat Disagree; 3 = Neither agree nor disagree; 4 = Somewhat agree; 5 = Strongly agree.

Dimension 1: Basic Attention (BA)

Items:

- BA1 (Q.1_1): I was fully absorbed in the VR activity, paying close attention to the tasks or scenarios presented
- BA2 (Q.1_2): My attention did not drift away from the VR experience to unrelated thoughts or external distractions
- BA3 (Q.1_3): I found it easy to concentrate on the VR experience for extended periods without becoming bored or tired
- BA4 (Q.1_4): The VR environment captured my attention immediately, keeping me engaged throughout

Dimension 2: Temporal Dissociation (TD)

Items:

- TD1 (Q2_1): While using the VR system, I completely lost sense of time passing in the real world
- TD2 (Q2_2): I was surprised by how much time had passed when I finished my VR session
- TD3 (Q2_3): My perception of time slowed down or sped up while I was immersed in the VR experience
- TD4 (Q2_4): I felt like time stood still during my engagement with the VR content

Dimension 3: Transportation (TRA)

Items:

- TRA1 (Q3_1): I felt as though I was truly 'inside' the VR world, experiencing its events first-hand
- TRA2 (Q3_2): The VR environment felt as real to me as the physical world
- TRA3 (Q3_3): I felt physically present in the virtual space, interacting with its elements as if they were real
- TRA4 (Q3_4): The boundary between the real world and the VR world blurred for me during the experience

Dimension 4: Emotional Involvement (EI)

Items:

- EI1 (Q4_1): The VR experience evoked strong emotions in me, such as excitement, joy, or empathy
- EI2 (Q4_2): I felt emotionally connected to the characters or story presented in the VR experience
- EI3 (Q4_3): The VR content triggered a deep emotional response that lingered even after the session ended

- EI4 (Q4_4): I was emotionally invested in the outcomes of the scenarios presented in the VR experience

Adaptation and Scoring

The local instrument retained four dimensions informed by Cheng and Tsai (2020) and used four locally worded items per dimension. Enjoyment, a further dimension in the source study, was not included. Items were rated on a five-point scale: 1 = strongly disagree, 2 = somewhat disagree, 3 = neither agree nor disagree, 4 = somewhat agree and 5 = strongly agree. The items above document the local adaptation rather than a verbatim reproduction of the source scale.

Each dimension score was the mean of its available item responses. Two participants omitted Q.1_2; no other item responses were missing. These scores were used descriptively, without a claim of psychometric validation for this cohort.

Source: Cheng and Tsai (2020); see the reference list.

Appendix B

IVR Usability Testing Checklist for Hazard Identification Training

This sheet is only for observers who are managing IVR training sessions to make sure whether they clearly prepare for every data collection step.

Note. This appendix reproduces the observer checklist. Its label “cognitive overload” was a prompt to note behavior, not a validated measure. Hazard scores and timing fields are not analyzed quantitatively in this report.

Participant Preparation

- ☐ Informing the goal of the usability test and interview.
- ☐ Participant briefed on IVR equipment use.

System Interaction and Navigation

- ☐Participant initiates IVR session without issues
- ☐Navigation controls understood and used effectively
- ☐Tools or options within IVR selected appropriately
- ☐Ease of movement within the IVR environment
- ☐Any navigation errors or difficulties recorded

Hazard Identification Tasks

- ☐Hazards correctly identified
- ☐Incorrect hazard identifications (false positives)
- ☐Hazards missed (false negatives)
- ☐Time taken to identify each hazard
- ☐Confidence level in hazard identification (if applicable, e.g., through post-task survey)

Usability and Experience

- ☐Instances of physical discomfort (e.g., dizziness, eye strain)
- ☐Signs of cognitive overload (e.g., frustration, asking for help)
- ☐Overall engagement with the IVR experience
- ☐Any technical issues encountered (e.g., lag, glitches)
- ☐Ease of exiting or pausing the IVR session

Feedback and Observations

- ☐Participant's verbal feedback on the IVR experience
- ☐Observer's notes on participant's behavior not covered by checklist
- ☐Suggestions for improvement (based on participant and observer feedback)

Appendix C

Definitions for the Retained Observational Evidence

Table C1

Operational definitions and limits of interpretation

Evidence category	Field or activity	Interpretation in this report
Interaction	Controller proficiency	Recorded categories: Low, Moderate, High and Excellent. Reported as counts and percentages, not a continuous score.
Interaction	Navigation difficulty	Recorded categories: None, Minor, Moderate, Significant and Major. These labels are not counts of errors; no numeric spacing is assumed.
Hazard activity	Identification and mitigation	Identification refers to locating or selecting a hazard. Mitigation is a subsequent safety response. No quantitative score is reported because first attempts, feedback and retries cannot be distinguished reliably in the archived summaries.
User experience	Difficulty managing task demands	Observer descriptions of confusion or requests for help. The checklist label cognitive overload does not establish a measured cognitive-load construct.
User experience	Physical discomfort	The 27 main-evaluation checklist records contain None. Pilot reports of dizziness inform design recommendations; no symptom-reduction effect is inferred.