# Advancing Usability of an Immersive Virtual Reality Team Training Environment within a Learning-Engineering Cycle

Parkhi Malhotra, Vipin Verma, Robert F. Siegle, Kamala Avancha, Kevin Gary, Jamie C. Gorman, Randall D. Spain, Benjamin Goldberg, & Scotty D. Craig

Immersive Virtual Reality | learning engineering | Usability Heuristics

*A Virtual Reality (VR) training environment for Tactical Combat Casualty Care (TC3) was built around the Team Dynamics Measurement Framework (TDMF) to address the need for adaptive team training inside a Casualty Collection Point (CCP). Guided by a Learning Engineering (LE) process, we embedded this work within nested LE cycles that emphasized iterative refinement. Two rounds of heuristic evaluation were conducted using a combination of VR and domain specific heuristics to assess the system's usability and its support for team training. This paper focuses on findings from the second heuristic evaluation which contribute to the design of scalable, team VR training aligned with LE principles.*

# Introduction

Today's military environments are increasingly dynamic, requiring teams to coordinate and perform complex tasks under time pressure and uncertainty. To address this challenge, our team developed a virtual reality (VR) training environment focused on Tactical Combat Casualty Care (TC3), for a casualty collection point (CCP) scenario that incorporates the Team Dynamics Measurement Framework (TDMF; Avancha et al., 2024) to develop and assess team adaptability skills (Craig et al., 2024). Virtual and synthetic environments have a long history of supporting effective training while reducing the risks and logistical burdens associated with live exercises (Andrews & Craig, 2015; Shubeck et al., 2016). However, VR based training systems present unique usability challenges such as inconsistent interactions, limited collaboration, and misalignment between physical and virtual worlds, which can hinder learning outcomes if left unaddressed (Sutcliffe & Gault, 2004; Derby et al., 2024).

To guide system development, we adopted a Learning Engineering (LE) approach, using nested LE cycles to integrate user-centred and domain specific feedback. A key part of this involved conducting heuristic evaluations to assess system usability. Heuristic evaluation is an expert driven usability testing method that identifies the mismatch between design and user expectations, using already established heuristics (Nielsen, 1994). Our first, heuristic evaluation, reported in Malhotra et al. (2025), applied a combination of heuristic for virtual environments (VE) proposed by Sutcliffe & Gault (2004) and the Derby Dozen principles introduced by Derby et al., (2024) using Nielsen's severity rating scale (Nielsen, 1992) to evaluate the impact of each identified usability issue, from minor to critical usability. These include issues in natural engagement and sense of presence, interaction with equipment, avatar height inconsistencies, system responsiveness, limited support for team-based interaction and orientation within the VR environment. These findings were fed back into the LE cycle as design requirements. The development team implemented a series of changes, including improved avatars, button interactions, medical equipment, refined navigation, orientation cues and adjustments to visuals where the key priority for the team was enabling multiplayer functionality to support team-based training. To evaluate the effectiveness of these changes, we conducted a second heuristic evaluation focused on identifying remaining usability concerns that could impact trainee performance.

# A Learning Engineering Process

This work describes our nested LE cycles within the development phase, focusing on iterative heuristic evaluations to inform development and incorporate TDMF to train team adaptability skills. The VR system and TDMF was developed following the LE process (Goodell & Kolodner, 2023; Kessler et al., 2023), which included four phases: (1) Challenge, where the problem is examined in context; (2) Creation, where solutions are designed to meet user needs; (3) Implementation, where the solution is deployed and data are gathered; and (4) Investigation, where the collected data are analyzed to evaluate the solution's impact. Nested LE cycles (Totino & Kessler, 2024; Craig et al., 2025) were used to (a) characterize the training challenge and end-user constraints through conversations with military partners and Subject Matter Experts (SMEs), (b) design training solutions using hybrid Cognitive Task Analysis (hCTA) and event-flow diagrams centered on patient care at a CCP and (c) specify perturbations that elicit adaptive team behavior while managing the CCP. Such nested cycles of design and evaluation are a common practice in LE, where smaller iterative loops within the larger phases help refine complex solutions (Avancha et al., 2024; Craig et al., 2025). SME reviews iteratively refined the scenario structure, timing, and scenario narrative so that the VR training would both align with doctrine and create opportunities to observe meaningful changes in team dynamics across successive perturbations.

# Method

The evaluation followed a structured protocol adapted from the first evaluation cycle (Malhotra et al., 2025) and heuristics relevant to the VR system were consolidated to assess six key dimensions of VE: Before entering the VR environment (Items=4); Navigation in VR (Items=4); Tasks within VR (Items=4); Feedback and Collaboration (Items=2); Post Task and Exit Scenario (Items=2); and Scenario-Specific (Items=1) (see Table 1). This was applied to the updated version of the VR training environment, which incorporated improvements to the avatar displays, interactions, role-based tasks, visual transitions, and

scenario flow. The purpose of the evaluation was to assess whether usability improvements made in response to the first-round findings effectively addressed previous concerns and to identify any new issues before implementing with trainees.

A team of three (n=3) human factors researchers conducted the evaluation, all of whom were familiar with immersive systems and had prior experience with the VR training scenario. Findings indicate that having expertise in usability evaluation, as well as in the specific interface domain being evaluated, leads to significantly better results. (Derby et al., 2024; Nielsen, 1994). The evaluation was conducted using Meta Quest 3 head-mounted displays. Each evaluator rotated through the three predefined trainee roles in the scenario: Prioritizer, Stabilizer, and Medical Supplier representing the structure of TC3 within the CCP setting. The evaluation included Scenario 1 and Scenario 2, with a 10-minute break in between sessions to minimize VR-related fatigue within a standard lab environment. During the sessions, each evaluator independently took notes on usability breakdowns, interface inconsistencies and confusing elements. The evaluators also had an option to think aloud while they were inside the VE and their observations were recorded. Evaluations were concluded individually to allow each participant to record their observations without influence from others. After completing their sessions, the evaluators convened to synthesize findings, discuss key usability issues, and assign consensus-based severity ratings.

Table 1.
Heuristic Checklist for VR Usability Evaluation

| Phase | Heuristic | Description |
|---|---|---|
| **Before Entering VR** | 1. Unboxing and Setup | Users should be introduced to the UI, features, and interaction methods. |
| | 1. Instructions | Instructions should provide actionable feedback. |
| | 1. Support for Learning | Cue active objects and provide explanations as needed. |
| | 1. Consistent Departures | Mark and apply design compromises consistently. |
| **Navigation in VR** | 1. Natural Engagement | Interaction should match real-world expectations. |
| | 1. Sense of Presence | The user should feel present in a real world. |
| | 1. Navigation and Orientation Support | Users should be able to locate themselves and reset positions. |
| | 1. Organization and Simplification | UI should focus on immersive elements, not external controls. |
| **Tasks within VR** | 1. Task Compatibility | Virtual tasks and object behavior should reflect real-world expectations. |

| | | |
|---|---|---|
| | 1. Natural Expression of Action | Allow physical actions; avoid restriction by hardware. |
| | 1. Integration of Physical and Virtual Worlds | Enable effective task completion using virtual tools. |
| | 1. Action and Representation Coordination | User actions and avatar behavior should align with <200ms delay. |
| **Feedback & Collaboration** | 1. Realistic Feedback | Actions should result in expected and immediate responses. |
| | 1. Team Collaboration | Include landmarks to orient users sharing virtual space. |
| **Post-Task & Exit** | 1. Comfort | Use should not cause discomfort or fatigue. |
| | 1. Clear Entry/Exit | Entry and exit should be intuitive and communicated. |
| **Scenario-Specific** | 1. Dynamic Adaptation | The system should respond to unexpected changes or disruptions. |

# Findings

The second heuristic evaluation identified a moderate number of usability issues, few of which persisted from the initial evaluation despite targeted design updates. Many previously critical problems such as interaction breakdowns, lack of avatar alignment, and successful implementation of multiplayer functionality were addressed. These updates resulted in improved presence, more reliable control and experience and support for role based changes.

However, evaluators ran into a few of lower-severity but instructionally relevant issues. E.g. scenario logic remained partially opaque: certain prompts, such as “Ask Location,” were inconsistently triggered. The scripts for dialogues by non playing characters (NPC) needed an update. This disrupted team coordination and created delays in both the scenarios. Other interface issues included overlapping menus, inconsistent exit wound toggle that hindered task navigation. Some object states such as crates or oxygen tanks did not update consistently across users, which limited shared situational awareness (SA).

Additionally, object behavior remained inconsistent in some cases: thermometers and nasal tubes occasionally clipped through patients or failed to respond to expected interactions, and critical equipment like medical bags (see Figure 1) spawned in inaccessible areas. This issue also highlights the limitations of testing inside a lab environment since spatial orientation remains a challenge, particularly when users removed and refitted headsets mid-scenario. These actions often caused avatar misalignment, floating hands. Instructions specifically have to be given to the participants by the experimenters, which should be a part of the scenario.

**Figure. 1.**
**Scene from the VR training environment showing available medical supplies**

Overall, the second evaluation reflected a notable reduction in the severity of usability concerns compared to the first round, but also emphasized the need for continued refinement in instructional support, feedback systems, and collaborative task alignment. These findings reinforce the importance of iterative evaluation during the LE cycle and provide actionable guidance as the project advances into the implementation phase with participants from the Army testing the system.

# Discussion

This study represents a nested LE cycle focused on refining a VR-based team training environment in a CCP setting. The initial evaluation identified critical usability issues in terms of Natural Engagement and Sense of Presence, including absence of user's avatar that informed substantial design changes. These included improvements to avatar height, interaction with objects, and multiplayer functionality. The second evaluation, presented in this paper, assessed the effectiveness of those changes and identified remaining usability concerns that, while less severe, could still affect the system's user experience, team training and coordination.

Findings from the second evaluation reaffirm the value of iterative, expert-driven usability evaluations in complex VR training systems within the LE cycle. Although the severity of issues decreased, gaps in design highlight the ongoing challenge of aligning system functionality with training objectives. These insights directly support the Investigation phase of the LE cycle

and will inform targeted refinements in interface behavior, role-based guidance, and scenario progression. The improvements from this cycle are expected to enhance user experience and the fidelity of analytics derived from user interactions.

The next step in the broader LE process is to transition into the Implementation phase. A formative evaluation with Army trainees will be conducted to assess the system's usability, effectiveness, and team training to generate ecologically valid insights on effectiveness, and usability in operational contexts. These findings will inform further refinement, guide the integration of Generalized Intelligent Framework for Tutoring (GIFT) based adaptive feedback, and support the development of a scalable, data-driven VR training solution grounded in LE principles.

# Acknowledgments

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research described herein has been sponsored by the U.S. Army Combat Capabilities Development Command under cooperative agreement W912CG-23-2-000. The statements and opinions expressed in this article do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

# References

Andrews, D. H., & Craig, S. D. (Eds.). (2015). Readings in training and simulation (Vol. 2): Research articles from 2000 to 2014. Human Factors and Ergonomics Society.

Avancha, K., Malhotra, P., Gorman, J. C., Verma, V., Spain, R., Goldberg, B., & Craig, S. D. (2024). Development of team dynamics measurement framework for adaptive teams. Proceedings of the 2024 Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC). National Training and Simulation Association.

Craig, S. D., Avancha, K., Malhotra, P., C., J., Verma, V., Likamwa, R., Gary, K., Spain, R., & Goldberg, B. (2025). Using a Nested Learning Engineering Methodology to Develop a Team Dynamic Measurement Framework for a Virtual Training Environment. In International Consortium for Innovation and Collaboration in Learning Engineering (ICICLE) 2024 Conference Proceedings: Solving for Complexity at Scale (pp. 115-132). https://doi.org/10.59668/2109.21735

Craig, S. D., Gary, K., Gorman, J. C., Verma, V., & LiKamWa, R. (2024). A synthetic training environment for assessing changes in team dynamics with the Generalized Intelligent Framework for Tutoring. In A. M. Sinatra (Ed.), Proceedings of the 12th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym12) (pp. 97–104). U.S. Army Combat Capabilities Development Command – Soldier Center.

Derby, S. K., Hughes, C. E., & Archer, R. D. (2024). The Derby Dozen: 12 usability heuristics for AR and MR. Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC). National Training and Simulation Association.

Goodell, J., & Kolodner, J. (Eds.). (2023). Learning engineering toolkit: Evidence-based practices from the learning sciences, instructional design, and beyond. Taylor & Francis.

Kessler, A., Craig, S. D., Goodell, J., Kurzweil, D., & Greenwald, S. W. (2023). Learning engineering is a process. In J. Goodell & J. Kolodner (Eds.), Learning engineering toolkit (pp. 29–46). Routledge.

Malhotra, P., Verma, V., Siegle, R. F., Avancha, K., Gorman, J. C., Spain, R. D., Goldberg, B. S., & Craig, S. D. (2025). Evaluating the usability of a VR team training environment utilizing extended reality (XR) heuristics. Proceedings of the Human Factors

and Ergonomics Society Annual Meeting, 69(1), 1729–1733. https://doi.org/10.1177/10711813251371033

Nielsen, J. (1992). Reliability of severity estimates for usability problems found by heuristic evaluation. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 129–130). ACM.

Nielsen, J. (1994). Enhancing the explanatory power of usability heuristics. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'94), 152–158. ACM.

Shubeck, K. T., Craig, S. D., & Hu, X. (2016). Live-action mass-casualty training and virtual world training: A comparison.Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 60(1), 2103–2107.

Sutcliffe, A. G., & Gault, B. (2004). Heuristic evaluation of virtual reality applications. Interacting with Computers, 16(4), 831–849.

Totino, L., & Kessler, A. (2024). "Why did we do that?" A systematic approach to tracking decisions in the design and iteration of learning experiences. The Journal of Applied Instructional Design, 13(2). https://doi.org/10.59668/1269.15630