ADVANCED MIXED REALITY TRAINING FOR FIRST RESPONDERS IN HURRICANE SCENARIOS

A Thesis
Presented to
The Faculty of the College of Graduate Studies
Lamar University

In Partial Fulfillment
of the Requirements for the Degree
Master of Engineering Science in Industrial Engineering
by
Muhammad Ilyas Mubarik
May 2026

ADVANCED MIXED REALITY TRAINING FOR FIRST RESPONDERS IN HURRICANE SCENARIOS

MUHAMMAD ILYAS MUBARIK

Approved:

Wenhao Yang
Supervising Professor

Yueqing Li
Committee Member

CHENG-HSiEN LIN
Committee Member

James Curry
Chair, Department of Industrial and System Engineering

Tracy Benson
Dean, College of Engineering

Jon Gray
Dean, College of Graduate Studies

ABSTRACT

ADVANCED MIXED REALITY TRAINING FOR FIRST RESPONDERS IN HURRICANE SCENARIOS

by

MUHAMMAD ILYAS MUBARIK

Extreme weather disasters, including hurricanes, necessitate that first responders operate in rapidly changing, high-risk environments requiring advanced situational awareness and decision-making abilities. Traditional training methods, such as live drills, classroom instruction, and tabletop simulations, are constrained by cost, logistical challenges, safety considerations, and limited environmental variability. This study proposes and evaluates an advanced Mixed Reality (MR)-based training framework that simulates hurricane-induced flood scenarios within an interactive and controllable environment.

The system is developed in Unity using the Universal Render Pipeline (URP) and deployed on the Meta Quest 3 headset. A custom depth-aware water shader with passthrough integration is implemented to achieve realistic flood visualization and environmental occlusion. The simulation incorporates parametric environmental controls, including dynamic water-level rise, rainfall intensity, storm lighting, and spatial audio, permitting controlled variation of scenario severity. Hand tracking and controller-based interaction methods allow users to perform rescue-related tasks within the simulated flood environment.

A structured experimental method is used to evaluate system effectiveness. Participants complete predefined response tasks under different environmental stress levels. Quantitative performance measures, including task execution time, decision accuracy, movement efficiency, and error frequency, are recorded and examined. Comparative scenario analysis is conducted to assess the impact of environmental intensity on cognitive and operational performance. Additionally, usability and perceived realism are measured applying standardized post-experiment questionnaires.

The platform supports rapid deployment, minimal setup time, and remote accessibility, facilitating scalable and distributed training sessions without the need for centralized infrastructure. Results indicate that the MR-based framework provides a controlled, repeatable, and cost-effective alternative to conventional disaster training methods, while enabling objective performance assessment.

This research advances the fields of emergency management and immersive simulation by integrating depth-aware environmental modeling, real-time parameter control, and quantitative evaluation within a scalable AR/MR training architecture.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my mentor and supervisor, Dr. Wenhao Yang, for his guidance, constant support, and kindness throughout my academic journey. Above all, I am especially thankful to him for rekindling my enthusiasm for engineering and for encouraging me throughout this work. I am deeply grateful to my thesis committee members, Dr. Yueqing Li and Dr. Cheng-Hsien Lin, for their valuable feedback, thoughtful suggestions, and continued support throughout my thesis defense and final revisions. This research was partially supported by the Lamar University Center for Resiliency Springboard Grant, for which I am sincerely appreciative.

I am deeply indebted to my parents, family, and friends back home for shaping the values, resilience, and strength that define me today. Their encouragement and belief in me have remained a constant source of motivation. I am equally grateful to my family here for their warmth, care, and comforting presence, which helped make the distance from home feel smaller. Finally, this work is dedicated to the memory of my grandparents, whose selfless sacrifices made the opportunities I pursue today possible. Their love, prayers, and legacy continue to inspire me every step of the way.

Table of Contents

List of Tables

List of Figures

Chapter I

Introduction

1.1 Global Context and Motivation

The frequency and severity of climate-induced disasters, including hurricanes, extreme flooding, and snow-related visibility hazards, have increased significantly in recent decades. These events generate complex, quickly evolving environments that require highly trained emergency responders capable of operating under physical, cognitive, and environmental stress. Flooding poses dynamic hazards, including increasing water levels, unstable terrain, debris movement, reduced visibility, and structural instability (Eskandarinejad et al. 2025). Effective response in such conditions depends not only on methodical knowledge but also on positional awareness, environmental cognition, and swift decision-making.

Traditional disaster training methods, classroom instruction, slide-based briefings, and periodic live drills have been widely adopted across emergency management agencies. (Gwynne et al. 2020) However, these approaches confront substantial limitations. Large-scale drills are logistically demanding, costly, difficult to repeat frequently, and often incapable of replicating progressive environmental escalation. Moreover, safety concerns restrict the realism of flood or extreme-weather simulations.

Recent research highlights that immersive technologies such as Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR) collectively referred to as Extended Reality (XR) offer promising alternatives for enhancing disaster preparedness and operational training (Khanal et al. 2022). XR systems allow the simulation of hazardous environments in controlled, repeatable, and scalable formats. These technologies can provide immersive experiences that support both cognitive and motor skill development without exposing responders to real-world risk.

1.2 XR in Disaster Management and First Responder Training

Over the past decade, XR technologies have increasingly been adopted in disaster management contexts (Feng and Gai 2026). A systematic review of XR applications in disaster management technology demonstrates growing implementation in emergency preparedness, hazard visualization, and risk communication (Khanal et al. 2022). These studies stress XR's capacity to improve situational understanding and mission readiness (Howard et al. 2020).

In first responder training specifically, immersive VR/AR platforms have shown measurable improvements in task accuracy and execution speed compared with traditional training approaches (Koutitas, Smith, and Lawrence 2021). Quantitative reviews suggest that AR/VR training frameworks can greatly improve response efficiency and familiarity with equipment and operational layouts. This suggests that immersive environments improve knowledge retention and operational recall (Xi et al. 2024).

Furthermore, AR-based training and command-and-control frameworks have proved the effectiveness of integrating sensor-based environmental overlays directly into responders' fields of view (Azpiroz et al. 2024). Such systems support real-time visualization of contextual data, enhancing mission coordination and environmental awareness.

Similarly, AR-based environmental hazard simulations such as snow visibility modeling for mountainous disaster preparedness have illustrated how immersive overlays can enhance risk perception and proactive decision-making (Saifi and Ramsankaran 2024). These findings support the broader hypothesis that immersive AR visualization improves environmental comprehension in hazardous conditions.

However, while these systems demonstrate success in equipment training, environmental visualization, and search scenarios, limited research addresses depth-aware flood simulation integrated within mixed reality environments.

1.3 Situational Awareness and Adaptive Interface Design

In emergency response operations, situational awareness is first important aspect in determining success. It involves the perception of environmental elements, understanding of their meaning, and the projection of future states (Toner 2010). Modern AR research focuses more on information delivery that must be user-friendly, context-sensitive, and role-dependent to avoid mental stress (Azpiroz et al. 2024).

Adaptive AR interface design focuses on modular systems in which information visibility to the user adjusts according to the kind of operation (e.g., survival-focused vs. search-focused tasks) and cognitive demands (Azpiroz et al. 2024). Overexposure to non-critical data may impair decision-making, whereas insufficient information may endanger safety (Misra, Roberts, and Rhodes 2020). Therefore, Environmental realism must be managed by XR systems with mental effort management.

This adaptive principle is particularly relevant in hurricane flood scenarios, where responders must continuously interpret environmental changes such as water level rise, structural occlusion, debris interaction, and reduced visibility. An effective mixed reality training system must therefore incorporate both environmental simulation and adaptive information filtering.

1.4 Representativeness and Skill Transfer

A major theoretical principle in immersive training design is representativeness, the degree to which simulated environments preserve perception–action couplings present within real-world contexts (Le Noury et al. 2023). Skill transfer can be reduced if sensory cues, motion timing, or object behavior differ noticeably from actual conditions. (Hasegawa, Okada, and Fujii 2021)

The hardware limitations evaluated by research for XR tools are frame rate constraints, limited haptic feedback, and field-of-view restrictions that may affect motion realism and timing precision (Le Noury et al. 2023). These findings emphasize the importance of engineered simulation environments, particularly when modeling dynamic physical phenomena such as water movement and occlusion.

For hurricane flood training, representativeness requires:

- Accurate depth perception and occlusion
- Realistic water interaction with physical structures
- Dynamic lighting and environmental visibility changes
- Interactive object behavior consistent with physical expectations

Without these features, immersive training may fail to replicate the complexity of real-world flood conditions.

1.5 Research Gap

Although prior XR studies have demonstrated improvements in operational familiarity, response efficiency, and environmental visualization, an important gap remains in mixed reality flood training for first responders. (TRACLabs 2018) Existing literature has not sufficiently addressed the development and evaluation of flood-focused MR training scenarios that combine environmental realism with real-world spatial integration. In particular, three limitations are especially relevant to this study.

- First, relatively few studies examine mixed reality flood-response training in which the participant experiences environmental hazards such as standing water, debris, degraded lighting, and storm-related audio within the same physical space (Tsujimoto, Fukuda, and Yabuki 2024).
- Second, limited work has focused on the integration of occluded real-world environment with virtual flood and elements, even though depth-aware occlusion is important for making floodwater appear grounded and interact with real environment in passthrough-based MR systems.
- Third, few studies provide a comparative evaluation of participant response in flood-intensified conditions and cognitive workload on the user within the same training framework.

Accordingly, the present research addresses a gap at the intersection of mixed reality flood training, real-world occlusion integration, and comparative evaluation of participant performance and perception under Normal AR and Flooded AR conditions.

1.6 Purpose of the Study

The purpose of this research is to design, implement, and experimentally evaluate a depth-integrated mixed reality hurricane flood training platform that:

- Simulates dynamic water-level rise using shader-based modeling
- Integrates real-world environment and occlusion

- Incorporates adaptive environmental controls (rain, lighting, sound)
- Enables interactive object manipulation through hand tracking
- Measures performance metrics under escalating environmental stress

The system is developed using Unity and deployed on a mixed reality head-mounted device capable of spatial mapping and depth sensing. Controlled experiments evaluate task completion time, error rate, movement efficiency, and decision accuracy across multiple environmental conditions.

1.7 Significance of the Study

This study contributes to:

- Industrial Engineering: by developing a quantitative performance evaluation framework for immersive disaster training
- Human–Technology Interaction: by integrating adaptive situational awareness principles into mixed reality systems (Azpiroz et al. 2024)
- Disaster Management: by extending XR applications into dynamic flood modeling contexts (Khanal et al. 2022; Saifi and Ramsankaran 2024)
- Immersive Training Research: by addressing representativeness constraints identified in XR studies (Le Noury et al. 2023)

By enabling rapid scheduling and remote deployment, the proposed system reduces reliance on centralized infrastructure and supports distributed emergency teams (Uhl et al. 2025).

Chapter II

Literature Review

2.1 Extended Reality in Disaster Preparedness and Response

In recent decade, research on extended reality (XR) in disaster management has expanded, particularly in applications involving emergency preparation, hazard communication, evacuation, and responder training. A systematic review by (Khanal et al. 2022) found that virtual reality (VR), augmented reality (AR), and mixed reality (MR) have been applied across different phases for training, to assess the infrastructure, public awareness, and predictive analysis of complex problems. (Zhu and Li 2021) similarly conclude that VR and AR have been increasingly adopted in emergency management in the built environment, with strong potential to enhance pre-emergency preparedness, response, and recovery.

The literature consistently highlights one practical advantage of XR over conventional drills: hazardous and rapidly changing conditions can be represented repeatedly without exposing trainees to actual risk. In hospital and emergency-preparedness contexts, (Jung 2022) and (Alshowair et al. 2024) report that virtual simulation can supplement tabletop or lecture-based training by improving repeatability, learner engagement, and realism. These findings are particularly relevant to flood-related training, where water movement, reduced visibility, and unstable surroundings are difficult to reproduce safely during routine instruction.

2.2 XR for First Responder Training

A good immersive system is required when training because trainees must take quick decisions while being aware of hazards, teammates, and the environment. (Koutitas, Smith, and Lawrence 2021) evaluated AR/VR training technologies for EMS first responders and reported that immersive training can improve performance while providing a controllable instructional environment. More recent work has continued to support the value of XR for emergency-service preparation, including training for mass-casualty and high-pressure scenarios (Heldring et al. 2024; Lochmannová et al. 2025).

Although many studies show promise for VR, fewer focus specifically on mixed-reality training, in which participants remain fixed in a real room while virtual hazards are layered into the space. This emerging direction is also reflected in mixed reality disaster-drill research, for example (Xu et al. 2023). Where the researcher demonstrated MR for indoor earthquake safety protocols, hazard environment, and emergency-response behavior. This difference is important for first responders because real-world mobility, depth judgment, and obstacle understanding are all required in the field performance (Coolen et al. 2020). In that regard, a stronger connection between environmental perception and action may be supported by mixed reality than by fully virtual training when carefully designed.

2.3 Situational Awareness and AR Interface Design

In emergency responses, situational awareness is fundamental since responders have to perceive relevant cues, interpret their significance, and predict how conditions may change. In AR and MR settings, this requirement creates a design challenge: the system is required to communicate useful information and hazards without overwhelming the user. (Azpiroz et al. 2024) argue that AR interfaces for first responders should be flexible, adjustable, and aligned with operational roles. Their work emphasizes that information delivery must support environmental understanding rather than distract from it.

Research related to AR for adverse-visibility response tasks similarly shows that the design of overlays, hazard markers, and visual information placement affects usability and mission support (Oregui et al. 2024). These findings are applicable to the present study because the flooded condition was not intended to function simply as a visual effect. Instead, the water, debris, dimming, and audio were designed to modify the participant's understanding of the environment and hence influence hazard interpretation and decision-making.

2.4 Environmental Realism, Representativeness, and Skill Transfer

A frequent theme in immersive training research is that realism alone is not sufficient; if learning is expected to transfer, the simulation must preserve meaningful perception-action relationships. The realism of space, hazards, timing, and sensory cues are included in disaster training. (Khanal et al. 2022) note that researchers must balance ecological validity against practical restrictions such as performance, usability, and simulator comfort. (Jung 2022) also reports that the practical usefulness of disaster simulation depends not only on submersion but on how well the scenario upholds meaningful training behavior.

For flood-related training, representativeness depends on whether the participant experiences the environment as operationally compromised. Standing water, hidden hazards, debris, degraded lighting, and storm-related sound all contribute to this perception. (Saifi and Ramsankaran 2024) shows that AR environmental simulation can be used to communicate hazardous conditions by visualizing degraded visibility. Their work, although focused on snow visibility, supports the broader idea that AR-based environmental overlays can be designed to represent dynamic hazard conditions in ways that support preparedness and interpretation.

A similar significance appears in mixed reality simulation research work. In the simulations design quality, it is evaluated not only by visual realism but also by how effectively the system delivers meaningful interaction, assessment, and training outcomes (Brunzini et al. 2022). This approach is relevant to my research where the simulation was not intended to display virtual flood but to create a meaningful environment in which users can experience the hazard, auditory conditions, and demonstrate measurable task performance.

2.5 Flood Visualization and Hazard Communication

Flood-specific visualization is a developing area within XR research. (Khanal et al. 2022) identify flood simulation and flood visualization as recurring but still limited topics within the broader XR disaster-management literature. (Zhu and Li 2021) likewise point to a continuing need for emergency-management applications that connect built-environment visualization with operational decision-making. These observations suggest that flood-focused MR training for first responders remains comparatively underdeveloped, especially when compared with more established XR applications in evacuation or hospital preparedness.

Communicating environmental change is another important issue. Alone water depth does not define flood conditions; spatial obstructions must also be considered by responders, reduced traction, uncertain footing, and the possibility that real objects or pathways are partially affected (Höllermann and Heidenreich 2025). This makes flood visualization in mixed reality more demanding than a simple overlay. (Rydvanskiy and Hedley 2021) Virtual flood must be perceived as part of the environment to contribute meaningfully rather than as an isolated visual layer (Sermet and Demir 2018).

2.6 Synthesis of Literature

Four conclusions supported by the literature directly shaped the present study. First, for disaster preparedness emergency-response training, XR is a credible and increasingly effective medium (Khanal et al. 2022; Jung 2022; Alshowair et al. 2024). Second, first responder applications benefit when XR systems strengthen situational and environmental understanding rather than only presenting static information (Azpiroz et al. 2024). Third, dangerous environment realism and representativeness matter because they impact whether the environment is interpreted as operationally meaningful (Zhu and Li 2021). Fourth, mixed reality flood-response training remains an underexplored area, especially the integration of environmental hazards into real space and evaluation across baseline and hazardous conditions.

Accordingly, this thesis builds on prior XR training literature while focusing on less-studied problem: the design and evaluation of a mixed reality flood-response scenario for first responders using Normal AR and Flooded AR conditions, depth-aware blocked floodwater, environmental management, and post-condition survey measures.

Chapter III

Methodology

3.1 System Design Overview

This chapter describes the hardware platform, software stack, scenario architecture, environmental implementation, user interface design, survey logic, and experimental procedure used to develop and evaluate the mixed reality hurricane-response training system. The purpose of this chapter is to document the implemented system in sufficient detail so that the design rationale, study flow, and technical decisions can be clearly understood and replicated.

3.1.1 Hardware and Software Setup

The mixed reality application was developed for the Meta Quest 3 headset and implemented in Unity as an immersive passthrough-based training system. The Quest 3 platform was selected because it supports mixed-reality passthrough, controller-based interaction, and environment-depth functionality, all of which were central to this project. The headset allowed the user to remain aware of the physical room while virtual flood hazards, survey interfaces, and control elements were overlaid in the scene.

Unity was used as the primary development environment because it provided real-time rendering, scene management, user interface tools, and scripting flexibility required to build the training application. The project was configured using the Universal Render Pipeline (URP), which provided a practical rendering framework for XR deployment, transparent materials, post-processing, and Shader Graph-based effects. URP was especially important for implementing virtual floodwater and maintaining a consistent rendering workflow across the mixed-reality scene.

The software stack also included the Meta XR SDK with passthrough and environment depth support. Passthrough was necessary because the training design required the user to remain visually connected to the real environment while experiencing virtual hazards. Environmental depth support was used to improve the spatial integration of virtual content, particularly the floodwater, by allowing real-world objects to occlude virtual surfaces where appropriate. This capability was essential for making the virtual flood appear physically grounded within the room rather than simply layered on top of the passthrough view (Meta 2025a; Meta 2025b).

Shader Graph was used to construct the visual water material. This approach provided flexibility in controlling transparency, appearance, and the flood surface's overall behavior. Rather than using a simple flat material, the water body was treated as a shader-driven component to better match the intended visual realism of the flooded condition. The use of Shader Graph also aligned with the project requirement of integrating occlusion-aware rendering behavior into the flood visualization (Meta 2025b; Meta 2025c).

Controller-based interaction and in-headset panel access were implemented using Meta Quest input support, while the world-space control menu and survey interface were implemented with XR-compatible canvas interaction. These interfaces relied on tracked-device UI interaction principles, enabling participants or researchers to access menu and survey elements within the headset environment (Meta 2026; Unity Technologies 2025a; Unity Technologies 2025b).

3.1.2 System Architecture

The application was organized as a modular mixed-reality training system comprising five major functional parts: the environmental state manager, the flooded-scene visual system, the researcher control interface, the survey subsystem, and the data logging subsystem. These components operated together to support both scenario delivery and structured data collection for studies.

The environmental state manager was responsible for enabling or disabling the two study conditions, namely the Normal AR environment and the Flooded AR environment. These conditions were represented as separate but related states so that the same base rescue setting could be shown either as a stable condition or as a hazard-intensified flood condition. This architecture helped maintain consistency in the spatial task setting while allowing controlled environmental manipulation.

The flooded-scene visual system included the floodwater object, debris hazards, dimming effect, audio cues, and other environmental manipulations associated with hurricane flooding. The researcher control interface was implemented as an in-headset menu that provided direct access to light dimming, water level, and sound intensity adjustment. The survey subsystem managed the condition-specific question blocks and the final mixed reality perception survey. The data-logging subsystem recorded participant responses, response times, and interaction-related measures for subsequent analysis.

The system's overall workflow followed a staged sequence. The participant first experienced one environmental condition, completed NASA-TLX externally, returned to the application to complete the corresponding question block, then repeated the same process for the second environmental condition. After both condition-related surveys were completed, the participant took a final mixed-reality perception survey. This architecture separated environmental exposure from questionnaire response while preserving a coherent within-subject study design.

3.2 Scenario Development

3.2.1 Base Rescue Environment

The scenario was developed as an indoor emergency-response space that could be presented under two environmental modes. The base rescue environment was kept intentionally consistent across study conditions so that participants would not face a fundamentally different task setting between the Normal AR and Flooded AR phases. This consistency was necessary to ensure that differences in response were attributable to environmental conditions rather than to changes in scene layout or general task structure.

The Normal AR condition represented the baseline state of the rescue environment. In this mode, the participant viewed the same operational space, with only minor flood-related hazards. The purpose of the normal condition was to provide a comparison point against which the hazard-intensified flooded condition could be evaluated.

3.2.2 Flooded Environment Construction

The Flooded AR condition was developed by adding several flood-specific environmental stressors to the base rescue scene. These included a visible water body, debris objects, reduced apparent lighting, and environmental audio. The purpose of this condition was to simulate a hurricane-affected operational environment in which responders must interpret hazards, maintain situational awareness, and operate in a compromised indoor space.

The flooded environment was not intended as a decorative variation of the normal scene. Instead, it was constructed as a hazard-intensified state that changed how the room appeared, how hazards were perceived, and how the participant interpreted the surrounding environment. This distinction was essential to the experimental design because the flooded condition served as the primary manipulated environment against which the baseline condition was compared.

3.2.3 Debris and Hazard Representation

Debris elements were placed throughout the flooded environment to represent the unstable and obstructed conditions typical of hurricane-related indoor flooding. These objects were used to visually communicate disorder, operational risk, and environmental compromise. Their inclusion increased the realism of the flood condition and strengthened the participants' perception that the environment was no longer fully safe or stable.

The debris also served an experimental purpose. By increasing clutter and visual hazard density, the flooded scene presented a more demanding interpretive environment than the baseline condition. This helped support the study's focus on hazard recognition, environmental interpretation, and condition-based comparison.

3.2.4 Floodwater Implementation

The floodwater was represented by a virtual water body placed along the floor region of the response environment. This water body was designed to simulate standing and rising water in an indoor rescue setting. The floodwater was positioned to visually occupy the participant's operating space and therefore function as both an environmental cue and a hazard indicator.

The water level could be adjusted through the in-headset environmental control menu. This made it possible to calibrate the apparent severity of flooding during scenario development, pilot testing, and controlled study setup. The purpose of this adjustable water level was to allow the researcher to vary the degree of environmental hazard while keeping the overall environment and task setting unchanged.

In addition to its visual role, the floodwater was treated conceptually as part of the environment's operational hazard structure. The participant was expected to interpret the presence of floodwater as a cue related to movement risk, environmental instability, and emergency severity.

3.2.5 Water Shader Graph and Occlusion Design

A shader-based approach was used to build floodwater rather than relying on a static colored surface. The water material was created in Shader Graph to customize transparency, visual blending, and the overall appearance of the flood surface. This was important for giving the water a more realistic and visually meaningful presence in the flooded condition.

The water body also required proper integration with the real-world environment. In mixed reality passthrough systems, transparent virtual surfaces can appear unrealistic if they render over real-world objects without regard to environmental depth. To address this issue, the water body was designed to work with environmental depth occlusion. This allowed real-world objects to visually occlude the water where appropriate, improving the realism of the flood scene and reducing the water's appearance as a simple overlay (Meta 2025a; Meta 2025b; Meta 2025c). as illustrated in figure 3.1.

The occlusion-aware design of the water body was particularly important to the study because it strengthened the perceptual realism of the flooded condition. By allowing physical furniture and room surfaces to visually interact with the floodwater, the system made the hazard appear more naturally embedded in the user's environment. This contributed directly to representativeness and environmental fidelity.

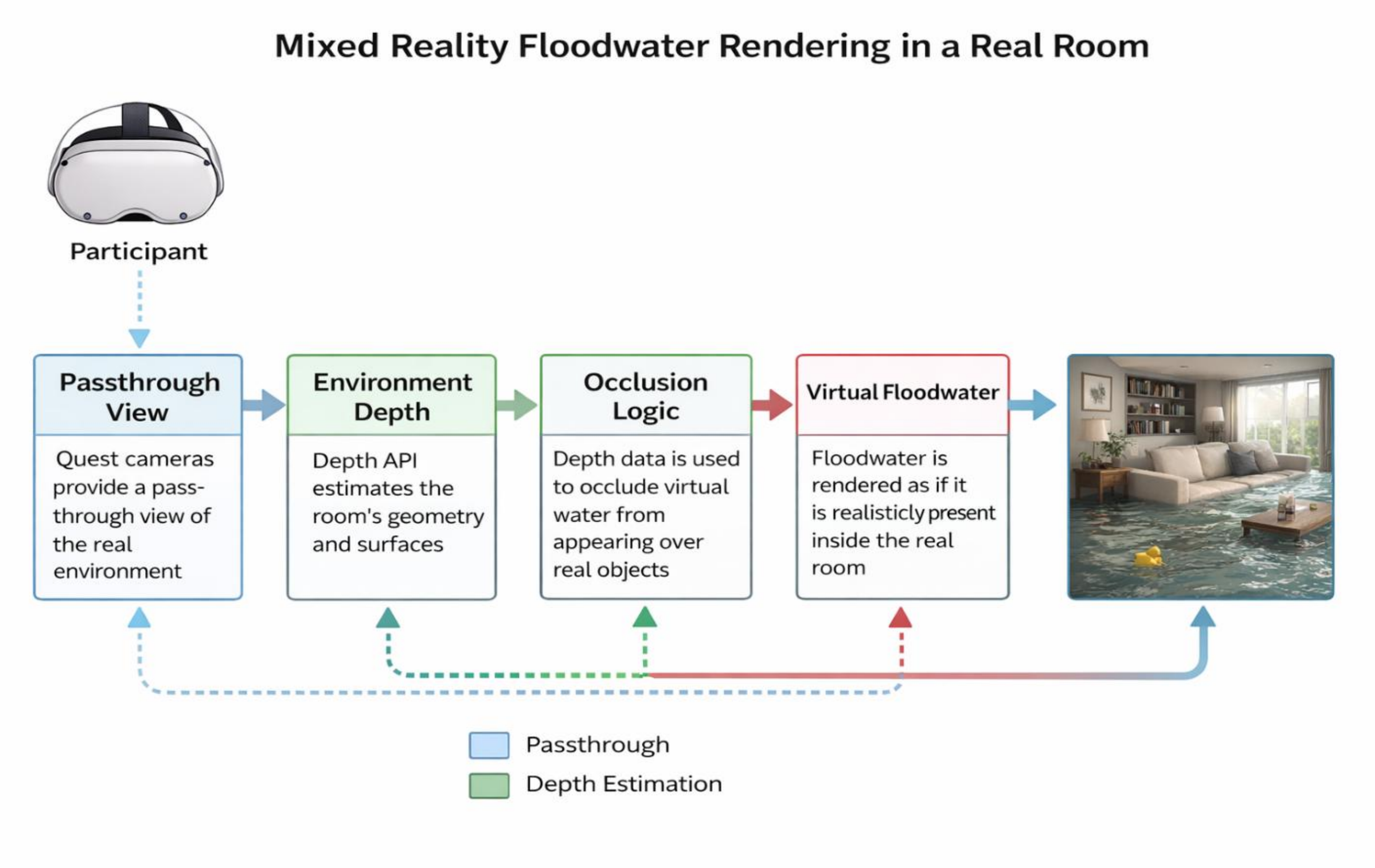

Figure 3.1 illustrates the conceptual operation of the occluded water body, including passthrough visualization, environment depth sensing, and the occlusion-aware floodwater material.

3.2.6 Audio Design

Environmental audio was incorporated to strengthen realism and support immersion. Three primary sound categories were used in the project: rain, wind, and floodwater. These sounds represented characteristic cues associated with hurricane and flood response conditions and were chosen to reinforce the sensory context of the flooded environment.

The purpose of the rain audio was to maintain the impression of an ongoing storm. Wind audio was used to communicate environmental instability and adverse weather conditions, while floodwater audio reinforced the presence and movement of water within the operational space. Together, these sounds created a more believable and stressful environmental backdrop than a silent scene.

The sound system was integrated with the researcher's control menu so that sound volume could be adjusted during testing and calibration. This allowed the researcher to control the strength of the auditory environment while maintaining the same general scenario structure.

3.2.7 Lighting Manipulation

A light-dimming function was implemented to simulate reduced visibility and disrupted environmental illumination. Because the application used mixed reality passthrough, simply reducing the light intensity of virtual objects was not sufficient to create a convincing perception of darkness. Instead, the system applied an overall dimming effect to reduce the scene's apparent brightness in the headset view.

This feature was included because poor visibility is a realistic component of hurricane-related emergency response. The dimming effect, therefore, served both a realism function and an experimental function by increasing the difficulty of environmental interpretation under degraded visual conditions.

3.3 Interface Design and Control Logic

3.3.1 Researcher Control Menu

An in-headset control menu was designed to support scenario setup and controlled manipulation of environmental variables. The menu was implemented as a world-space interface so that it could be accessed within the MR scene without returning to the Unity editor during runtime. This menu was primarily intended for researchers rather than for participant-directed task interfaces.

The menu included controls for water level, light dimming, and sound volume. Each control had a direct operational purpose. Water-level adjustment enabled flood severity to be calibrated. Light dimming allowed visibility degradation to be introduced or modified. Sound volume control allowed the intensity of the storm-related audio environment to be adjusted. Together, these menu functions provided a practical mechanism for scenario calibration and controlled environmental setup.

3.3.2 Survey Interface

The questionnaire subsystem was implemented as a world-space canvas displayed within the mixed-reality environment. The survey interface showed one question at a time, together with a set of selectable answer options. Toggle-based answer selection was used for both the multiple-choice scenario questions and the five-point Likert items in the final mixed reality perception survey.

The survey logic dynamically updated the question text, answer options, and section labels during runtime. The interface was designed to progress sequentially, recording one answer at a time before loading the next item. This design reduced clutter and supported a focused response experience in the headset.

3.3.3 Condition Selection Logic

The study interface used explicit condition selection so that the participant could return to the application after each environmental exposure and begin the corresponding question block. A participant identifier and environment selection state were used to determine whether the system should present the first condition-specific question set, the second remaining question set, or the final MR survey.

This logic was important because NASA-TLX was completed externally rather than in the headset. As a result, the software design needed to support repeated re-entry into the survey workflow while preserving the correct order of question presentation and preventing question overlap between the two environmental conditions.

3.4 Data Collection Methodology

3.4.1 Research Design

The study used a within-subject design in which each participant experienced both environmental conditions. One condition represented a baseline state, and the other represented a flood-intensified state. The within-subject structure was selected because it allowed the same participant to serve as their own comparison across the two environmental contexts, reducing the effect of between-participant variation.

To reduce question-related learning effects, the study did not use one fixed set of questions for each condition. Instead, the system used a shared bank of ten scenario questions. For each participant, five questions were assigned to the first completed condition, and the remaining five were assigned to the second condition. This ensured that the two environmental blocks used non-overlapping question sets while still drawing from the same shared question bank.

3.4.2 Participant Procedure

The study procedure began with participants' exposure to a single environmental condition in mixed reality. During this phase, the participant experienced either the Normal AR or the Flooded AR scenario. After this environmental experience, NASA-TLX was completed externally rather than inside the headset.

The participant then returned to the MR application and completed the five assigned scenario questions for that condition. After the first condition-specific question block was completed, the participant experienced the second environment, again completed NASA-TLX externally, and then returned to answer the remaining five scenario questions. After both condition-specific blocks were completed, the participant was presented with the final mixed-reality perception survey.

This procedure separated environmental exposure from questionnaire response while maintaining a controlled study sequence. It also ensured that workload assessment was collected externally and that the in-headset interface was reserved for the condition-related and perception-related surveys.

3.4.3 Data Collection

The system collected multiple types of data relevant to participant performance and perception. These included scenario-question responses, answer correctness on condition-specific knowledge items, response time, and post-experience mixed-reality perception ratings. NASA-TLX was collected separately, outside the headset, and therefore served as an external workload measure associated with each environmental condition.

In addition to answering data, the software architecture supported interaction logging related to user interface selection behavior. This design enabled recording not only what participants answered but also how they interacted with the study interface during the experiment.

3.4.4 Final Mixed Reality Survey

After both environmental condition surveys were completed, the system presented a final mixed reality survey using a five-point Likert response format. This survey was designed to capture participants' perceptions of display clarity, comfort, realism, immersion, situational awareness, and the overall usefulness of the system as a training supplement.

The final MR survey served a different purpose from the scenario questions. Whereas the condition-specific questions focused on flood-response interpretation and operational knowledge, the final MR survey focused on participants' experience with the training system itself.

3.5 Summary

The system described in this chapter combined a mixed reality rescue environment, a hazard-intensified flood condition, Shader Graph-based floodwater, depth-aware occlusion, debris hazards, environmental audio, light dimming, a researcher control menu, and a structured survey subsystem. The methodology combined two environmental exposures, external NASA-TLX administration, condition-specific question blocks, and a final mixed reality perception survey. Together, these elements formed the technical and methodological foundation of the thesis investigation.

Chapter IV

Data Collection

This chapter describes the experimental design, participant flow, data sources, logging architecture, data cleaning steps, and the statistical procedures used to analyze the collected data. The study evaluated first-responder performance and perceived workload across two mixed reality conditions: a normal augmented reality environment (Condition A) and a flooded augmented reality environment (Condition B).

4.1 Research Design

A within-subjects repeated-measures design was used. Each participant completed both environmental conditions to enable direct comparisons between normal and flooded scenarios while reducing inter-participant variability. To control sequence effects, participants were assigned to one of two presentation orders: A-B or B-A. In the A-B sequence, participants completed the normal condition first followed by the flooded condition. In the B-A sequence, participants completed the flooded condition first followed by the normal condition.

The in-headset assessment used the same objective question bank in both conditions. Five randomized multiple-choice questions were drawn from Bank B for each condition. After completing a condition, participants exited the headset and completed the NASA-TLX externally. After both task conditions were completed, the researcher conducted a short interview outside the headset.

4.2 Participants and Sample Structure

The analyzed dataset contained 16 complete participant records. According to the session summary log, 8 participants completed the A-B sequence, and 8 participants completed the B-A sequence. For the in-Quest assessment, each participant answered 5 questions in the normal condition and 5 questions in the flooded condition, resulting in 160 total response-level records. See table 4.1 and 4.2.

Table 4.1. *Analytical Datasets Used in the Study.*

Source	Rows	Structure	Key variables
NASA-TLX dataset	32 rows	16 participants x 2 conditions	Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, Frustration, Raw Score, Weighted Score
In-Quest response dataset	165 rows	5 Bank-B questions per condition, plus system records	Question ID, selected answer, correct answer, correctness, response time
Session summary dataset	16 rows	1 row per participant	Sequence group, total shown, total correct, accuracy, average response time

Table 4.2. *Sample Overview and Assessment Structure*

Statistic	Value
Participants (n)	16
A-B sequence	8
B-A sequence	8
Questions per condition	5
Total in-Quest responses	160

4.3 Instruments and Measures

Three data sources were used in the analysis.

NASA-TLX forms were collected externally after each task condition. Each participant contributed one normal-condition NASA-TLX record and one flooded-condition NASA-TLX record, producing 32 workload records in total.

In-Quest response logs produced by the mixed reality assessment application. These logs contained participant ID, condition label, question ID, selected option, correct option, correctness, response time, and timestamps.

Session summary logs are generated after the task is completed. These files captured the presentation order (A-B or B-A), the number of questions shown, the overall number of correct responses, the percentage accuracy, and the average response time.

The primary subjective outcomes were the six NASA-TLX subscales, the raw NASA-TLX score, and the weighted NASA-TLX score. The primary objective outcomes were question accuracy and mean response time in the mixed reality task.

The interaction log also recorded UI interactions. However, the final Quest interaction pipeline did not consistently provide reliable button-hit coordinates for every selection. Therefore, spatial precision fields that could not be recovered were retained as unavailable values in the raw logs and were not treated as primary inferential outcomes.

4.4 Experimental Procedure

At the beginning of the session, the participants entered or confirmed their identifier and selected the study order in the Start Panel. The selected order determined whether the normal environment or the flooded environment appeared first.

Condition A (Normal AR): participants experienced the normal augmented reality environment and answered 5 randomized Bank B questions in-headset.

External NASA-TLX (Task 1): immediately after the first condition, participants completed NASA-TLX outside the headset.

Condition B (Flooded AR): participants experienced the flooded augmented reality environment and answered a different set of 5 randomized Bank B questions in-headset.

External NASA-TLX (Task 2): immediately after the second condition, participants completed NASA-TLX outside the headset.

Interview: after both conditions were complete, the researcher conducted the post-task interview.

4.5 Data Collection and Logging Pipeline

The Quest application exported question-level responses and session summaries as comma-separated value (CSV) files. The external NASA-TLX data were compiled into a separate CSV file. The response log stored the following variables for each answered question: participant ID, condition, question ID, question text, selected option, correct option, binary correctness, response time, and timestamps. The session summary log stored one row per participant and included the sequence order and aggregated performance values.

The NASA-TLX file stored one row per participant per condition. Each row included Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, Frustration, Raw Score, and Weighted Score. Because each participant contributed both a normal-condition and a flooded-condition NASA-TLX record, workload comparisons could be analyzed as paired observations.

4.6 Data Cleaning and Preparation

Before analysis, the three CSV datasets were reviewed and cleaned. First, participant identifiers were checked across all files to ensure that the NASA-TLX file, response log, and session summary file referred to the same participants. Second, condition labels were harmonized to a common Normal-versus-Flooded structure. Third, response-level records were checked for completeness so that each participant contributed five responses in the normal condition and five responses in the flooded condition. Fourth, the session summary file was used to recover the sequence assignment (A-B or B-A) for order-effect analysis.

For inferential analysis, condition-level accuracy and mean response time were aggregated per participant to enable paired Normal and Flooded comparisons. NASA-TLX values were also paired by participants. Interaction-log precision fields that were unavailable in the final Quest pipeline were not used as primary analytic variables. This decision avoided introducing false precision values into the inferential analysis.

No records were removed from the final analysis set. The cleaned analytical sample, therefore, consisted of 16 participants, 32 NASA-TLX rows, 160 response rows, and 16 session summary rows.

4.7 Statistical Analysis

Descriptive statistics were computed for all major dependent measures. For NASA-TLX, means and standard deviations were computed for each subscale and for the raw and weighted scores under the Normal and Flooded conditions. For the in-Quest assessment, condition-level accuracy and mean response time were computed per participant, then summarized across participants.

Because each participant experienced both conditions, paired-sample comparisons were used to evaluate overall differences between the Normal and Flooded conditions. Wilcoxon signed-rank tests were applied to NASA-TLX dimensions, weighted score, participant-level accuracy, and participant-level mean response time. To examine sequence effects, the data were additionally summarized by order group (A-B and B-A) and then compared descriptively within each order.

The alpha level was interpreted conventionally at 0.05. Given the modest sample size, p-values slightly above 0.05 were treated as directional trends rather than conclusive effects.

Chapter V

Results and Discussion

This chapter presents descriptive and inferential results from the NASA-TLX dataset, the mixed-reality question-response dataset, and the sequence-order summaries. Results are presented first for the overall Normal-Versus-Flooded comparison and then by sequence group to show how the A-B and B-A orders influenced the observed pattern of outcomes.

5.1 Data Overview

The final paired dataset included 16 participants who completed both conditions. Each participant contributed one Normal NASA-TLX record and one Flooded NASA-TLX record. Each participant also completed ten in-Quest objective items in total: five in the Normal condition and five in the Flooded condition.

At the descriptive level, the NASA-TLX data suggested a higher workload in the Flooded condition for Mental Demand, Physical Demand, Temporal Demand, and Effort. The question-response data suggested a modest reduction in accuracy under the Flooded condition, whereas response time was nearly unchanged.

5.2 Overall NASA-TLX Results

Table 5.1. *Overall NASA-TLX Comparison between Normal and Flooded Conditions.*

Metric	Normal Mean	Flooded Mean	Mean Difference (Flooded-Normal)	Wilcoxon p
Mental Demand	28.75	38.75	10.00	0.025
Physical Demand	25.62	32.50	6.88	0.073
Temporal Demand	19.69	30.62	10.94	0.016
Performance	40.94	31.56	-9.38	0.108
Effort	35.62	47.81	12.19	0.002
Frustration	26.88	30.00	3.12	0.343
Raw Score	29.28	35.52	6.24	0.011
Weighted Score	29.34	38.02	8.68	0.001

The Wilcoxon signed-rank test results indicate that several subscales reached or approached statistical significance, suggesting that flood exposure meaningfully increased perceived task demand across multiple dimensions.

Table 5.1 shows that the Flooded condition has made higher mean workload scores than the Normal condition across most NASA-TLX dimensions. The most eye-catching statistically significant effects were observed for Weighted Score ($p = 0.001$) and Effort ($p = 0.002$), both of which surpassed the conventional significance threshold of $p < 0.05$. as illustrated in figure 5.1 and 5.2, the mean difference among the normal and flooded condition in weighted and raw score is 8.68 and 6.24 respectively These numbers indicate that users experienced significant greater overall workload. and they exerted noticeable effort in the flood condition as compared to the normal condition. Temporal Demand also reached statistical significance ($p = 0.016$), which shows the participants' perceived increase in time pressure during flood exposure. Mental Demand similarly crossed the significance threshold (p = 0.025), suggesting that flooded conditions imposed a greater cognitive load compared to normal conditions.

There is an increase in mean score from 25.62 to 32.50 in Physical demand, but this difference did not reach statistical significance (p = 0.073>0.05). Frustration exhibited the smallest mean difference (3.12) and the lowest level of statistical evidence (p = 0.343), suggesting that participants' affective responses did not differ substantially between conditions.

Point to be noted here, Perceived Performance moved in the inverse direction, with participants reporting lower self-assessed performance under flooded conditions (31.56 vs. 40.94). Although this difference did not reach statistical significance (p = 0.108 > 0.05) but there is a strong directional trend with the broader pattern of increased workload.

Taken together, these findings indicate that flooded environment conditions impose a significantly higher overall workload burden, particularly with respect to effort, temporal pressure, and cognitive demand, compared to normal MR conditions. The overall score of both conditions is illustrated in figure 5.3.

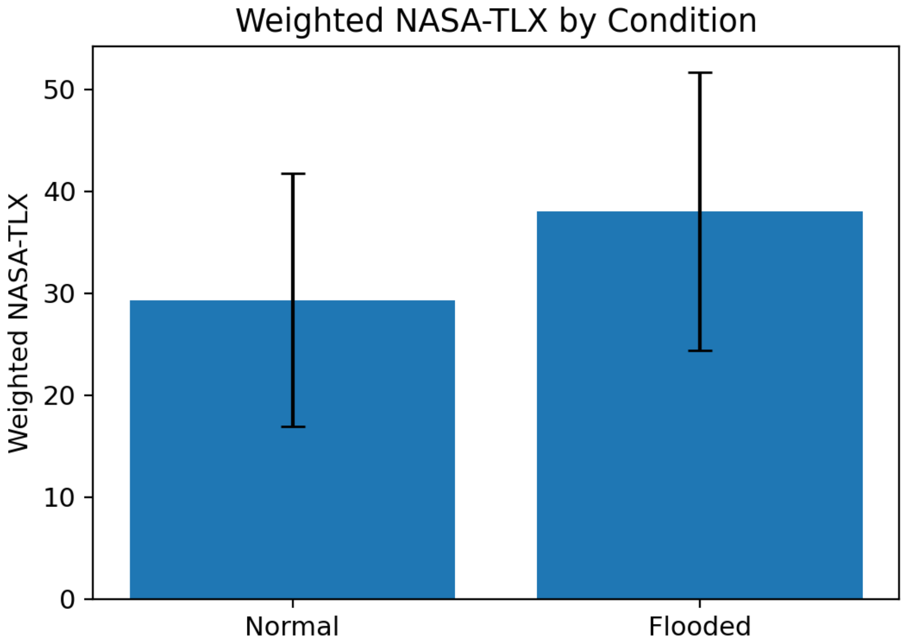

Figure 5.1. Overall Weighted NASA-TLX by Condition

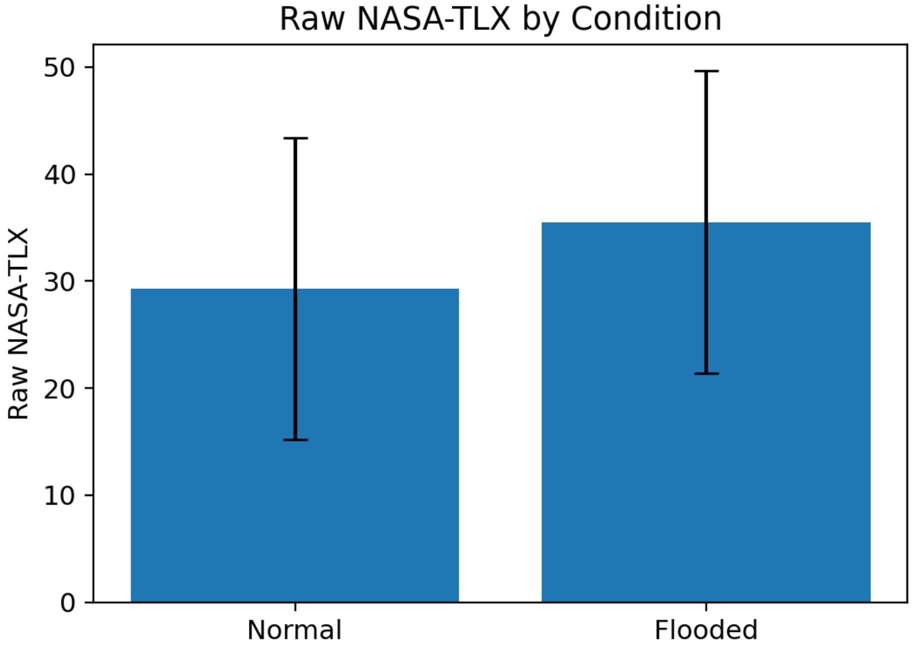

Figure 5.2. Overall Raw NASA-TLX by Condition.

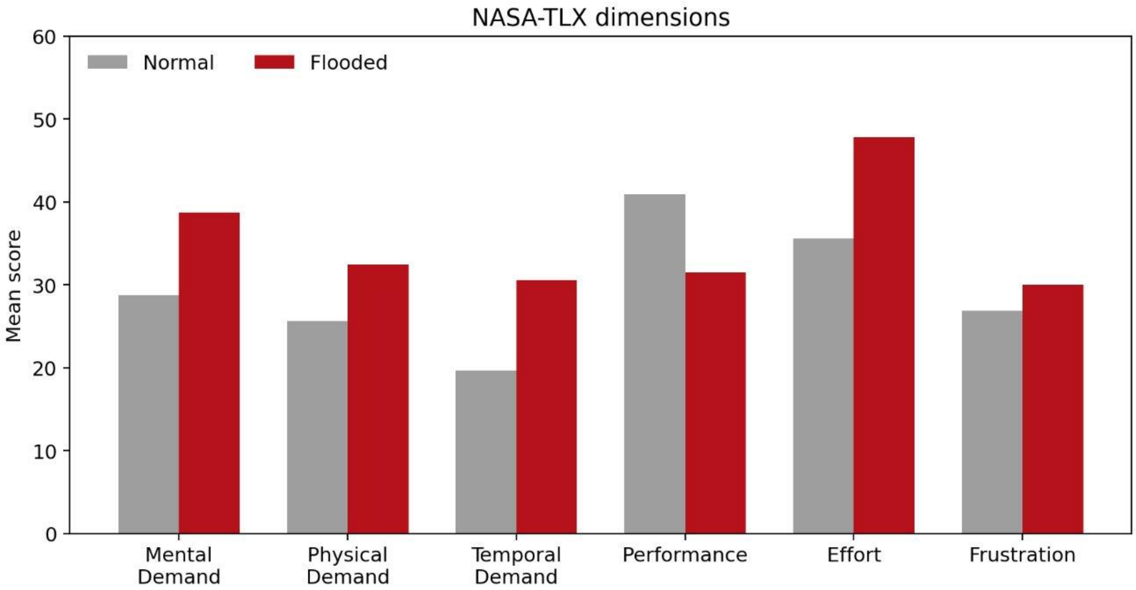

Figure 5.3. Key NASA-TLX Dimensions by Condition.

5.3 Overall Objective Performance Results

Table 5.2 shows that objective task performance was modestly lower in the Flooded condition than in the Normal condition. For the in-Quest question responses, the Normal condition produced higher mean accuracy than the Flooded condition (87.50% vs 80.00%). See figure 5.4. Mean response time was also lower in the Normal condition (18.06 s) than in the Flooded condition (21.11 s). see figure 5.5. However, neither the accuracy comparison nor the response-time comparison reached statistical significance. The accuracy result approached significance ($p = 0.070$), there is a possibility that performance-related effect will be clear in larger study, whereas response time showed weaker evidence ($p = 0.464$). These findings suggest that the flooded environment increased subjective workload more clearly than it changed objective task outcomes in this sample.

Table 5.2. *Overall in-Quest performance Comparison Between Normal and Flooded Conditions.*

Metric	Normal Mean	Flooded Mean	Mean Difference (Flooded-Normal)	Wilcoxon p
Accuracy (%)	87.50	80.00	-7.50	0.070
Response Time (s)	18.06	21.11	3.05	0.464

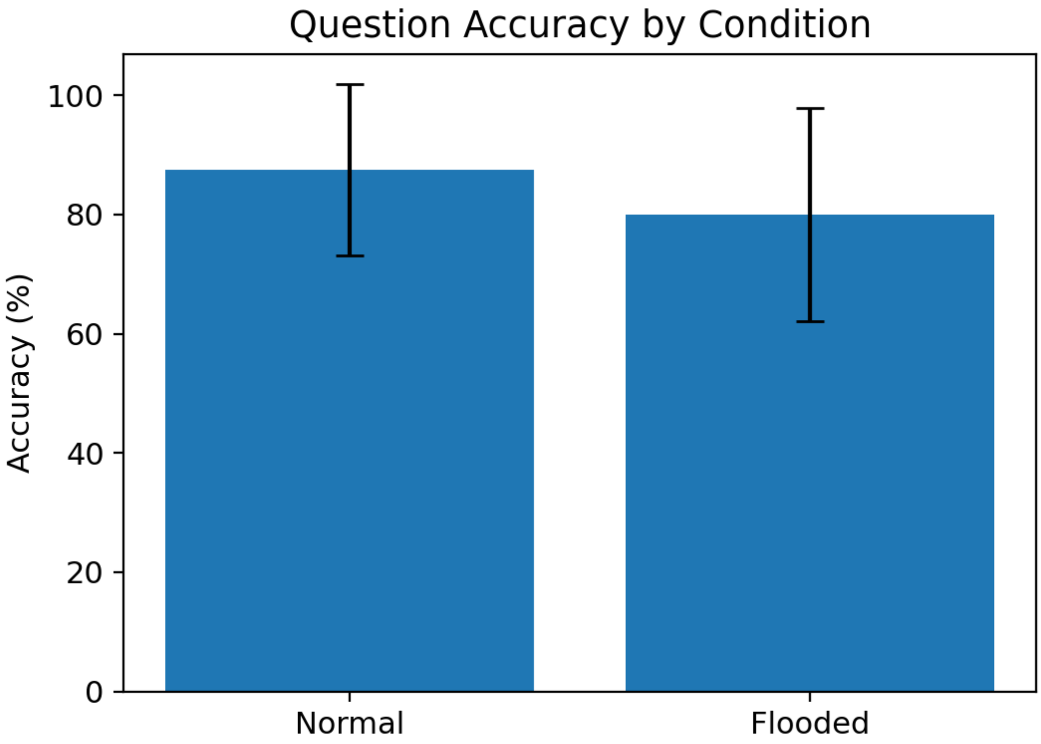

Figure 5.4. Mean Participant Accuracy by Condition.

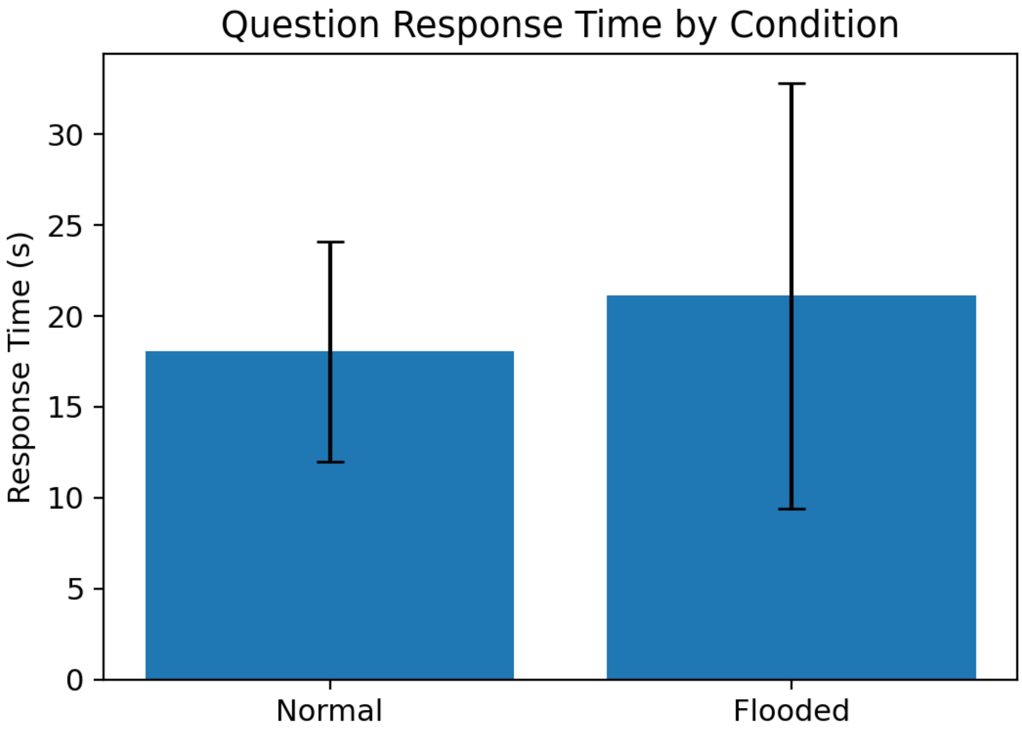

Figure 5.5. Mean Participant Response Time by Condition.

5.4 Sequence (A-B vs. B-A) Analysis

To examine possible order effects, the cleaned dataset was separated into the A-B group and the B-A group. The sequence analysis shows that the Flooded condition remained more demanding in both groups, but the magnitude of the difference varied by order. In the A-B group, the workload increase in the Flooded condition was more pronounced, particularly for Mental Demand and Effort. In the B-A group, the workload pattern remained directionally consistent, though the differences were attenuated across several dimensions.

Table 5.3. *Sequence-Specific NASA-TLX Comparison by Condition*

Order	Metric	Normal Mean	Flooded Mean	Mean Difference (Flooded-Normal)	Wilcoxon p
A-B	Mental Demand	24.00	36.50	12.50	0.043
A-B	Physical Demand	22.50	33.50	11.00	0.062
A-B	Temporal Demand	17.00	28.00	11.00	0.125
A-B	Performance	33.50	34.00	0.500	1.00
A-B	Effort	35.00	49.50	14.50	0.008
A-B	Frustration	28.00	33.00	5.00	0.344
A-B	Raw Score	26.67	35.75	9.08	0.010
A-B	Weighted Score	26.70	37.63	10.93	0.004
B-A	Mental Demand	36.67	42.50	5.83	0.406
B-A	Physical Demand	30.83	30.83	0.000	1.00
B-A	Temporal Demand	24.17	35.00	10.83	0.031
B-A	Performance	53.33	27.50	-25.83	0.031
B-A	Effort	36.67	45.00	8.33	0.125
B-A	Frustration	25.00	25.00	0.000	1.00
B-A	Raw Score	33.64	35.14	1.50	0.688
B-A	Weighted Score	33.75	38.67	4.92	0.219

The sequence-specific workload results indicate that the A-B group demonstrated a clearer and more statistically robust increase in workload from Normal to Flooded conditions than the B-A group. See table 5.3, In the A-B group, the Weighted NASA-TLX score increased from 26.70 in the Normal condition to 37.63 in the Flooded condition (p = 0.004), while Mental Demand increased from 24.00 to 36.50 (p = 0.043), both reaching statistical significance. In the B-A group, the Weighted NASA-TLX score increased from 33.75 in the Normal condition to 38.67 in the Flooded condition see figure 5.6; however, this difference did not reach statistical significance (p = 0.219).

This pattern suggests that the impact of the flooded environment on individuals' workload was more feasible when participants encountered the normal condition first. A reasonable interpretation is that before exposure to the normal condition, a baseline was established against which the demands of the flooded condition were perceived more deeply. Breaking down the difference between the two approaches, the flooded condition has stronger workload ratings in the A-B group.

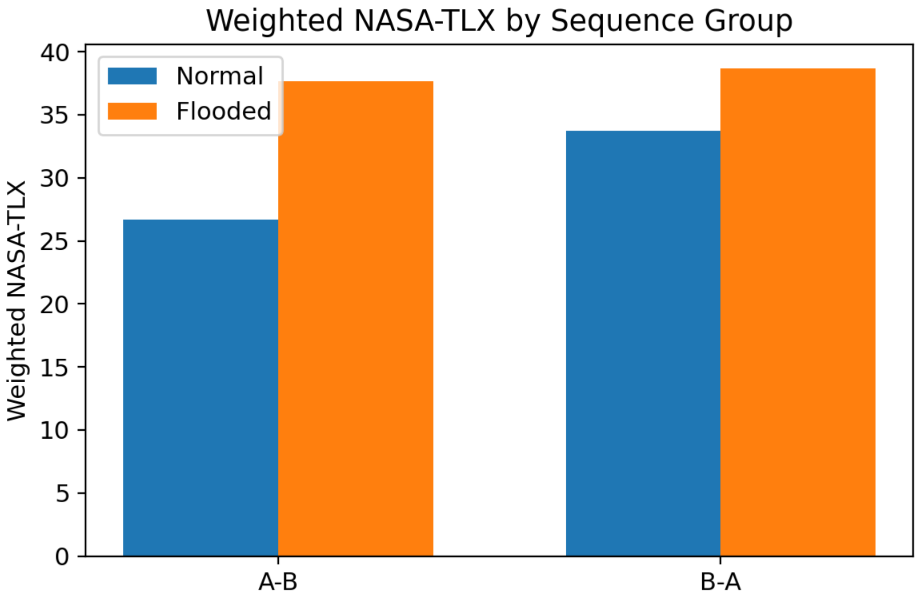

Figure 5.6. Weighted NASA-TLX by Sequence and Condition.

Table 5.4. *In-Quest Performance Results by Sequence Group*

Order	Metric	Normal Mean	Flooded Mean	Mean Difference (Flooded-Normal)	Wilcoxon p
A-B	Accuracy (%)	88.00	84.00	-4.00	0.531
A-B	Response Time (s)	16.13	21.25	5.12	0.322
B-A	Accuracy (%)	86.67	73.33	-13.33	0.250
B-A	Response Time (s)	21.27	20.86	-0.405	0.844

Table 5.4 summarizes objective performance by sequence group. In the A-B group, accuracy was slightly higher in the Normal condition than in the Flooded condition (88.00% vs. 84.00%), In the B-A group, the difference in accuracy was larger, decreasing from 86.67% in the Normal condition to 73.33% in the Flooded condition.

Response time showed a small sequence-dependent pattern as well, increased from 16.13 s to 21.25 s. In the A-B group, average response time was slightly higher in Flooded than in Normal, whereas in the B-A group Normal response time was slightly longer than Flooded as shown in see figure 5.8. Although these differences were not statistically significant in the current sample, they indicate that sequence may have influenced how participants adapted to task demands. See figure 5.7.

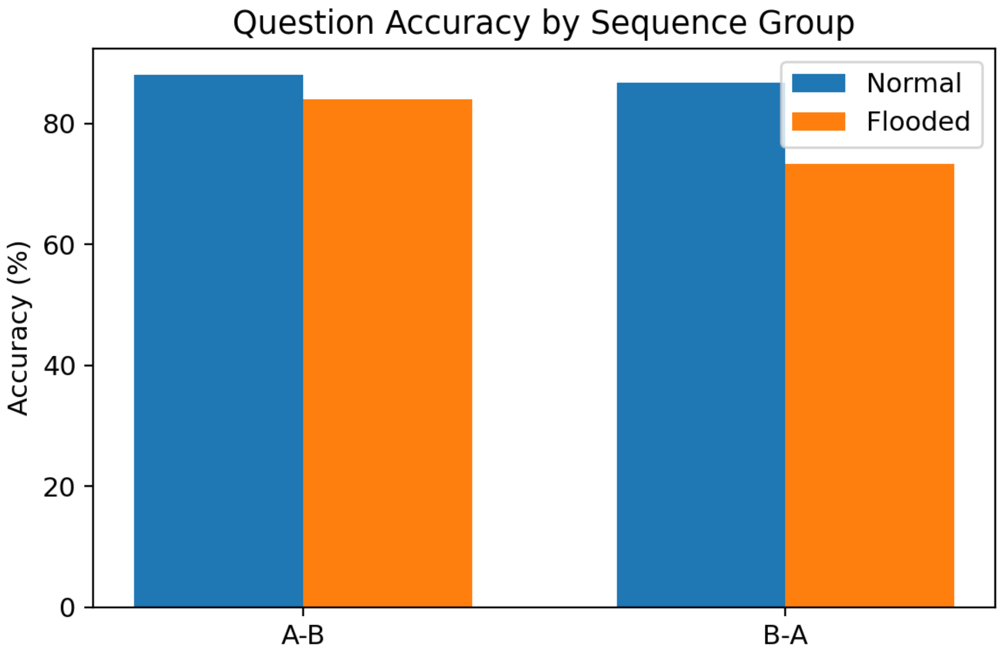

Figure 5.7. Accuracy by Sequence and Condition.

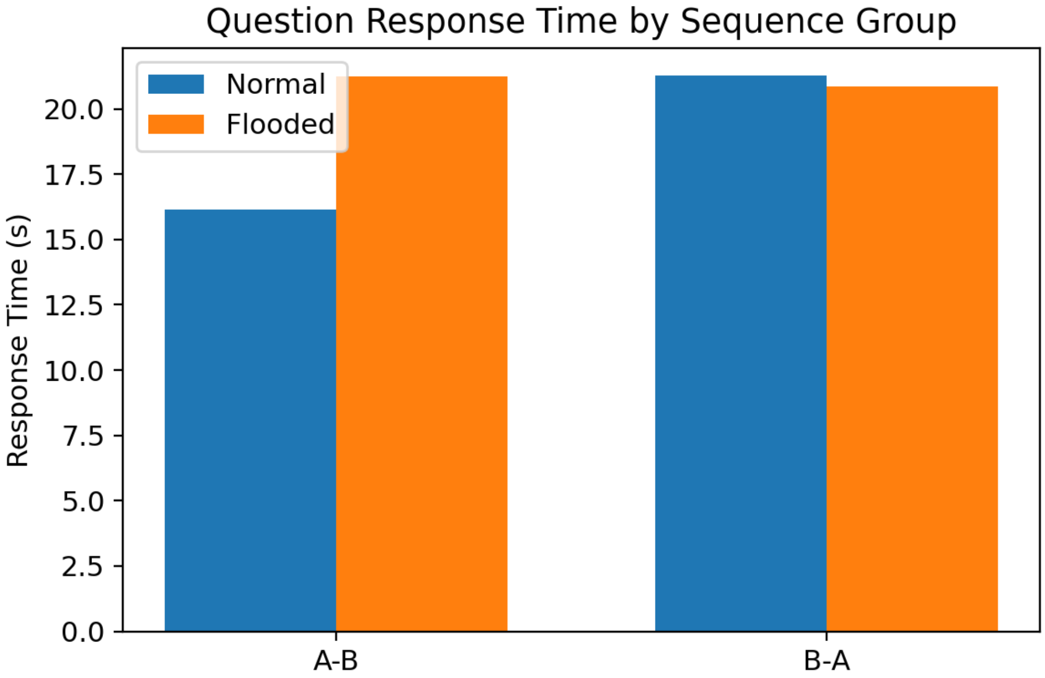

Figure 5.8. Response Time by Sequence and Condition.

5.5 Participant-Level Trends

Participant-level trends in weighted NASA-TLX revealed that most participants did not respond uniformly to the change in condition. Several participants showed substantial increases in weighted workload from Normal to Flooded, whereas others showed relatively little change. This spread is consistent with the moderate between-subject variability observed in the standard deviations as shown in figure 5.9.

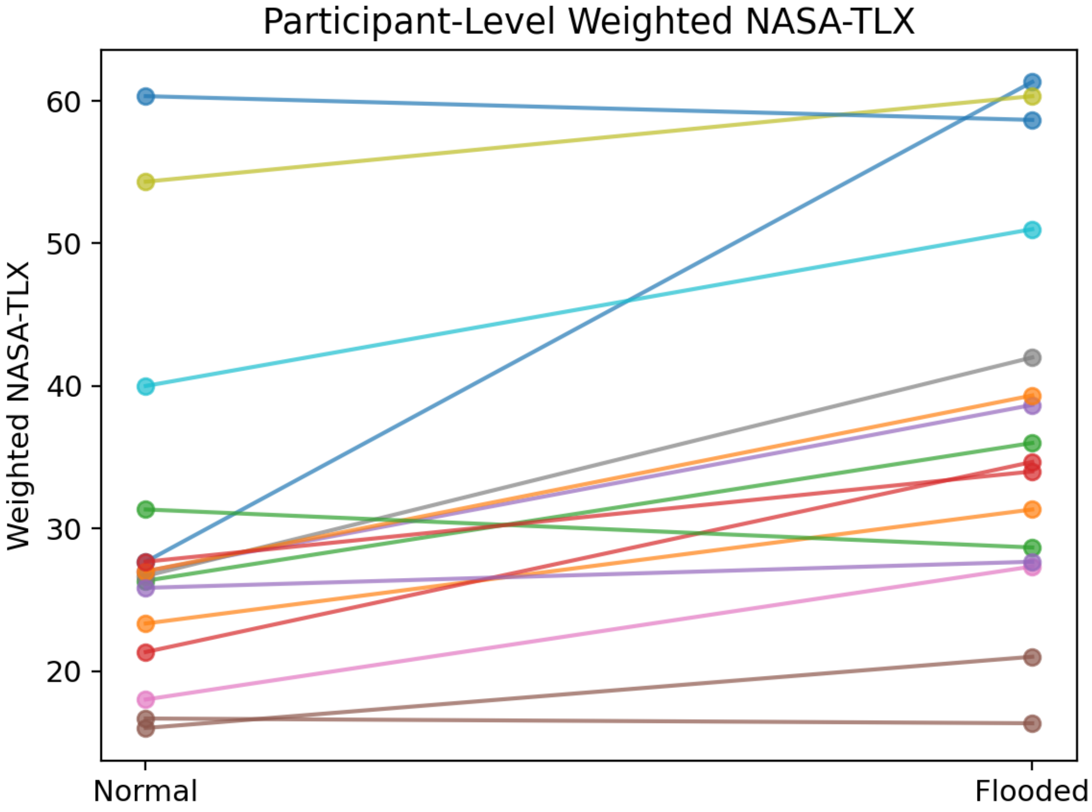

Figure 5.9. Participant-Level Change in Weighted NASA-TLX from Normal to Flooded.

5.6 Closed-Ended Question Results

The closed-ended responses show that participants generally viewed the AR flood simulation positively. The strongest results appeared in relation to the believability of the water dynamics, improved understanding of water behavior, and recommendation of the tool for first responder education. More mixed responses were observed for comfort, display clarity, and visual authenticity, which suggests that the system shows promise but still requires technical and visual refinement. Figure 5.10 presents the percentage distribution of responses for each closed-ended survey item.

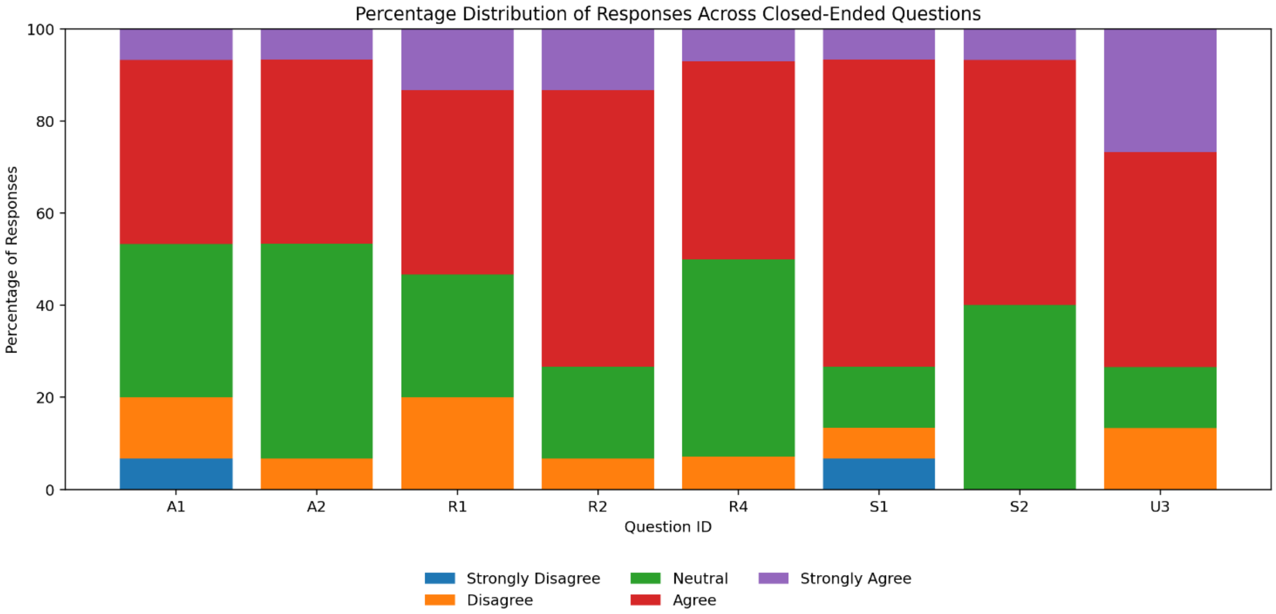

Figure 5.10. Percentage Distribution of Responses Across Closed-Ended Survey Questions.

Figure 5.11 groups each question into positive (Agree + Strongly Agree), neutral, and negative (Disagree + Strongly Disagree) response categories.

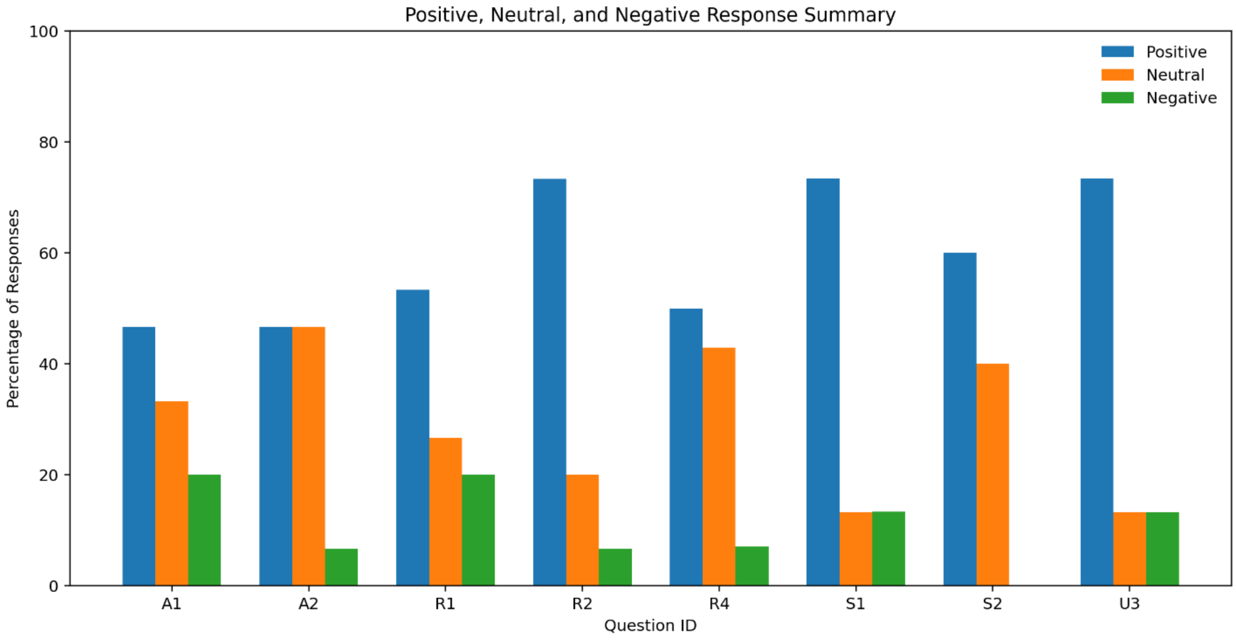

Figure 5.11. Positive, Neutral, and Negative Response Summary for Closed-Ended Questions.

5.7 Open-Ended Question Summary

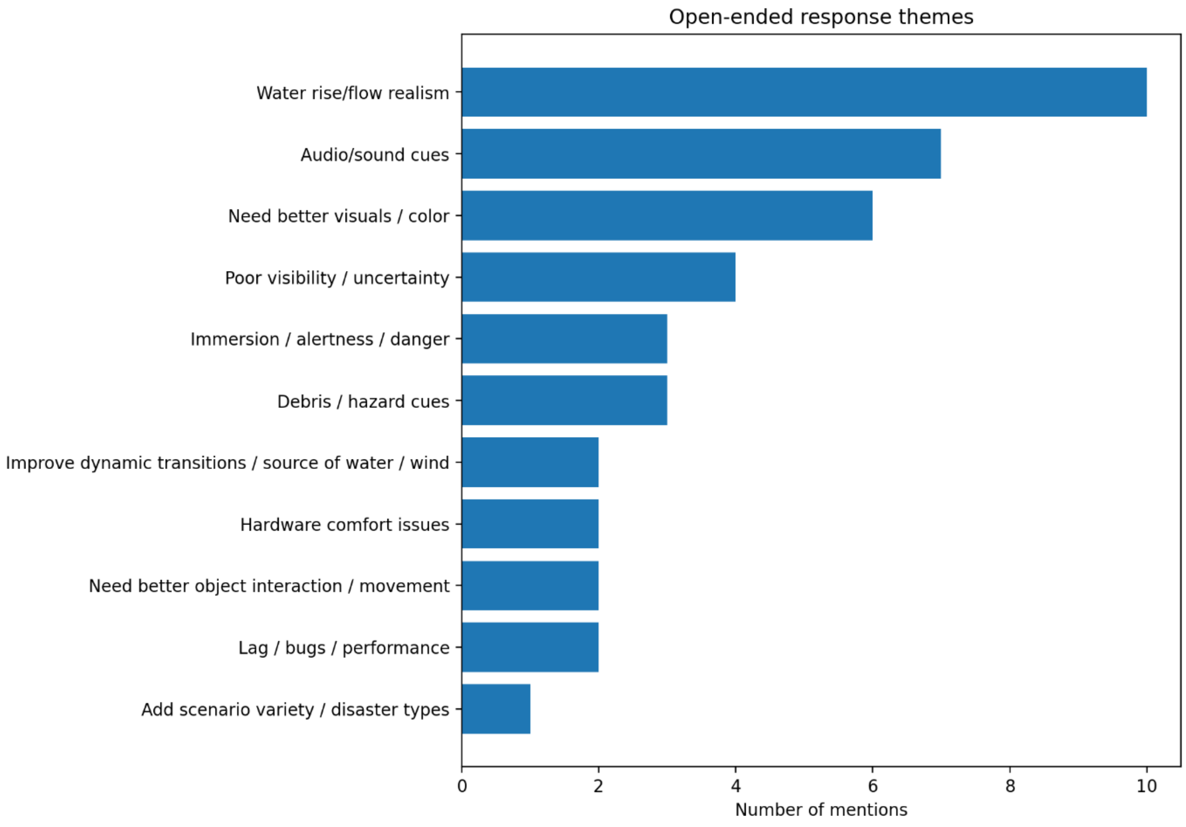

Figure 5.12. Frequency of themes identified in the open-ended responses.

Taken together, the survey results suggest that the mixed reality system was successful in communicating flood behavior and creating a meaningful sense of environmental risk. The strongest aspects of the experience were the dynamic water behavior, audio cues, and debris-related realism, all of which contributed to immersion and improved user understanding see figure 5.12. However, the survey also showed that comfort, visual refinement, and system smoothness remain important challenges. This combination of positive educational value and identifiable technical limitations indicates that the system is promising as a supplemental training tool but would benefit from additional development before broader deployment.

Overall, the open-ended responses indicate that the most effective components of the simulation were the water-related behaviors, environmental sounds, debris cues, and reduced visibility, all of which strengthened realism and hazard perception. At the same time, participants consistently identified the need for better visual detail, smoother interaction, improved performance, and broader scenario coverage. These findings support the conclusion that the AR flood simulation has strong training potential, especially as a tool for immersive hazard awareness, while also highlighting the improvements needed to make the experience more realistic, stable, and effective.

5.8 Summary of Findings

Taken together, the results indicate that the flooded mixed reality environment increased perceived workload more reliably than it changed raw task speed. The strongest overall directional effects were observed for Mental Demand, Temporal Demand, Effort, and perceived Performance. Accuracy declined modestly in the Flooded condition, but the change was smaller than the shift observed in workload measures.

The sequence analysis further suggests that order mattered. The A-B group demonstrated a clearer increase in subjective workload in the Flooded condition, while the B-A group showed a larger accuracy gap between Flooded and Normal. These sequence-dependent patterns support the decision to counterbalance presentation order and should be acknowledged in the interpretation of the findings.

Overall, the dataset supports the conclusion that the flooded environment imposed additional subjective burden on participants and introduced a modest performance cost, particularly when considered alongside the order in which the two environments were presented.

Chapter VI

Conclusion

This thesis presented the design and evaluation of an advanced mixed reality training system for first responders in hurricane-related flood scenarios. The purpose of the study was to investigate whether a mixed reality environment enriched with floodwater visualization, debris, degraded lighting, and environmental audio could provide a more realistic and operationally meaningful training context than a baseline augmented reality condition. The system was implemented on Meta Quest 3 using Unity, the Universal Render Pipeline, and depth-aware occlusion techniques to improve the spatial realism of virtual floodwater within a real-world environment.

The findings indicate that the flooded mixed reality condition imposed a greater subjective workload on participants than the normal condition. In particular, mental demand, temporal demand, effort, and overall NASA-TLX scores increased in the flooded scenario, showing that the hazard-rich environment was experienced as more cognitively demanding. This result is important because the goal of the system was not simply to display a flood visually, but to create an operational context in which participants had to interpret and respond to a more challenging environment. In that sense, the flooded condition functioned as intended by increasing perceived task demand and environmental pressure.

The objective performance results showed a smaller but still meaningful pattern. Participant accuracy declined modestly in the flooded condition, while response time changed only slightly. This suggests that the primary effect of the flooded environment was not necessarily a dramatic slowdown in response, but rather an increase in subjective burden combined with a moderate performance cost. The sequence analysis further showed that presentation order influenced the observed pattern of results. Participants who experienced the normal condition first tended to show a clearer increase in workload when later exposed to the flooded condition, while other order-related differences appeared in the accuracy results. These findings support the value of the counterbalanced design and indicate that adaptation and contrast effects likely influenced how participants experienced the two environments.

The perception survey and open-ended feedback further strengthen the conclusions of the study. Participants generally responded positively to the mixed reality simulation, particularly with respect to the believability of water behavior, the usefulness of the experience for understanding flood hazards, and the overall potential of the system as a training aid. The open-ended responses showed that dynamic water behavior, environmental sound, debris cues, and reduced visibility were among the most effective features for creating immersion and a sense of risk. At the same time, participants consistently identified several areas requiring improvement, including headset comfort, smoother interaction, stronger visual refinement, and more stable technical performance. These comments indicate that the system already demonstrates educational and experiential value, but that further development is needed before broader implementation.

Overall, this research supports the conclusion that mixed reality has strong potential as a supplemental training approach for first responders in flood-related disaster scenarios. By integrating virtual hazards into real space, the system provides a controllable, repeatable way to simulate environmental stress without the cost, safety risks, or logistical challenges of live disaster drills. The study also contributes to the broader literature on immersive training by showing that depth-aware flood visualization, environmental manipulation, and structured performance evaluation can be integrated into a single mixed-reality framework. Although the present system remains developmental and the sample size was modest, the findings demonstrate that hazard-rich mixed reality training can meaningfully influence perceived workload, environmental interpretation, and user engagement. In this regard, the thesis establishes a practical foundation for future work on mixed-reality disaster training systems that are more realistic, more scalable, and more directly aligned with first-responder preparation.

6.1 Limitations:

This study used general volunteers from a university as test subjects. The sample size consists of 16 participants. They were not professional first responders like firefighters or other emergency medical responders. This greatly minimizes the applicability of the findings since users conducted the workload testing without domain expertise. Although the mixed reality simulations test replicated a real scenario but there was no direct evaluation from domain experts. Because of the lack of expert evaluation, it was difficult for us to judge how well the simulation fulfilling the training requirement. This experiment tested only one condition, which is hurricane-flood scenario. And every user must go through same environment and procedure. Despite that fact environmental parameters were flexible, such as light, flood level, noise level etc. still some of the results from users were biased. The collected sample size is enough to apply the Wilcoxon signed-rank test. But keeping in mind that, this study lacks a priori power to test any effect or medium effects between scenario conditions in question. Comparing the data involving user accuracy and response time in the normal and flooded condition did not yield significant difference. Hence there is presence of some noticeable trends. I also want to conclude that the results of this study do not justify the other emergency scenarios like fire, casualties, chemical spillage etc.

NASA-TLX assessment in measuring the workload is another limit that needs to be addressed. This assessment is validated on self-report scale and data varying on individual interpreting rating. Yet, participant workload is affected by various factors, including but not limited to biased response, the way they feel about the training environment etc. not every user was meant to be a first responder or think like them, so that affect reflected in the study survey in the form of user interest, engagement, and curiosity to learn about emerging technology. In the study survey, all the Instructions to the participant were provided verbally about usability of the device by the researcher. This gave us the risk of unnecessary variability in the process due to the subjectivity involved by each user. For example, some participants might have difficulty understanding the survey steps due to their technical skills deficiencies and understanding. There was no familiarity workshop organized to teach the user about the use of meta quest device and hand controllers which caused significant variation in workload and assessment response.

6.2 Future Work:

Starting with the meta quest device or any other HMD, making the user familiar with the device, controllers, hand gestures and its practicality is first step in implementation of future work and it is recommended to organize a workshop to train the participants on device usage before starting the study. Along with that, more performance and testing parameters are required to simulate in the environment, which gives a tougher scenario to the user. It is recommended to give or teach the user about organized structured procedure of the study and no other information is provided during the testing, which will help to get unique, unbiased, workload, and assessment data. The most important direction is to conduct a structured case study with actual first responder trainees or professional emergency personnel, such as firefighters, paramedics, flood rescue operators, and emergency management officers it will help to evaluate the effectiveness of the training simulation and learning of the trainee. This study involved general participants from different domains of life, which was appropriate for an initial evaluation of the system, but it does not provide enough operational feedback. The lack of quality data makes it difficult to determine whether the simulated environment, task demands, and decision requirements truly reflect professional emergency educational and training practice. Participants with hurricane-specific training would be able to judge whether the flood environment feels realistic and the task involved in the scenario aligns with actual response priorities and procedures, and whether the overall training environment is strong enough for meaningful use in preparedness training anf education. Their feedback would provide a more grounded assessment of the system's practical relevance and would help identify which parts of the simulation need improvement.

A second direction for future work is to improve the evaluation and training procedure itself so that testing can be carried out more consistently among participants. In the current study, as told earlier that some guidance was provided during the session to help participants move through the environment and complete the required steps. Otherwise, the participants were unable to complete any step. for the development stage, we can consider it useful but future work should rely on more clearly structured operation and steps. which includes but not limited to participant onboarding, headset familiarization, a short practice phase, and a consistent sequence of instructions built directly into the simulation or delivered properly before the training and examination. This would reduce variation caused by unfamiliarity with the interface and would allow later studies to focus more clearly on the participant's response to the environment rather than on the process of learning how to use the system.

Future work should also expand both the measurement approach and the scope of the training system. Along with accuracy, response time, and workload ratings, further studies should include more detailed behavioral measures such as hesitation time, repeated actions, missed or useless trigger button presses, and interaction behavior within the simulated environment. At the same time, the current prototype should be extended beyond a single flood scenario so that the framework can be tested in other emergency contexts, such as dangerous construction scenarios, chemical handling, etc., which helps to increase the training demands. Long-term research is also important to figure out that whether repeated use of the simulation system improves learning, preparedness, and retention over time. Together, these steps would help move the platform from early research simulations toward a more credible, solid, and practically useful training tool.

References

Alshowair, Abdulmajeed, Jean Bail, Fatima AlSuwailem, Asmaa Mostafa, and Amro Abdel-Azeem. 2024. “Use of Virtual Reality Exercises in Disaster Preparedness Training: A Scoping Review.” *SAGE Open Medicine* 12: 20503121241241936. https://doi.org/10.1177/20503121241241936.

Azpiroz, Izar, Igor García Olaizola, Xabier Oregui, Anaida Fernández García, Verónica Ruiz, Blanca Larraga-García, and Álvaro Gutiérrez. 2024. “White Paper on Adaptive Situational Awareness Enhancing Augmented Reality Interface Design on First Responders in Rescue Tasks.” *Applied Sciences* 14 (18): 8282. https://doi.org/10.3390/app14188282.

Brunzini, Agnese, Alessandra Papetti, Daniele Messi, and Michele Germani. 2022. "A Comprehensive Method to Design and Assess Mixed Reality Simulations." *Virtual Reality* 26 (4): 1257–1275. https://doi.org/10.1007/s10055-022-00632-8.

Coolen, Bert, Peter J. Beek, Daphne J. Geerse, and Melvyn Roerdink. 2020. "Avoiding 3D Obstacles in Mixed Reality: Does It Differ from Negotiating Real Obstacles?" *Sensors* 20 (4): 1095. https://doi.org/10.3390/s20041095.

Eskandarinejad, Alireza, Rouzbeh Nazari, Mohammad Reza Nikoo, David Arellano, Shahram Pezeshk, and Seyed Hooman Ghasemi. 2025. “A Comprehensive Review of Geotechnical Implications of Floods and Water-Driven Disasters.” *Science of the Total Environment* 985: 179731. https://doi.org/10.1016/j.scitotenv.2025.179731.

Feng, Xinhang, and Wenmei Gai. 2026. "Extended Reality Technology Applied to Emergency Evacuation Research, Training and On-Site Guidance: Analysis and Future Research."

Reliability Engineering & System Safety 268: 112022. https://doi.org/10.1016/j.ress.2025.112022.

Gwynne, Steve, Martyn Amos, Max Kinateder, Noureddine Bénichou, Karen Boyce, C. Natalie van der Wal, and Enrico Ronchi. 2020. "The Future of Evacuation Drills: Assessing and Enhancing Evacuee Performance." *Safety Science* 129: 104767. https://doi.org/10.1016/j.ssci.2020.104767.

Hasegawa, Yumiko, Ayako Okada, and Keisuke Fujii. 2021. “Skill Differences in a Discrete Motor Task Emerging From the Environmental Perception Phase.” *Frontiers in Psychology* 12. https://doi.org/10.3389/fpsyg.2021.697914.

Heldring, Sara, Marie Jirwe, Jan Wihlborg, Lena Berg, and Veronica Lindström. 2024. “Using High-Fidelity Virtual Reality for Mass-Casualty Incident Training by First Responders—A Systematic Review of the Literature.” *Prehospital and Disaster Medicine* 39. https://doi.org/10.1017/S1049023X24000049.

Höllermann, Britta, and Anna Heidenreich. 2025. "Certainly Uncertain: The Role of Uncertainty Perception for Flood Risk Preparedness and Response." *Natural Hazards* 121 (18): 22027–22048. https://doi.org/10.1007/s11069-025-07675-5.

Howard, James P., II, Arthur O. Tucker, IV, Stephen A. Bailey, and James L. Dean. 2020. “Mixed Reality for Post-Disaster Situational Awareness.” *Johns Hopkins APL Technical Digest* 35 (3). https://secwww.jhuapl.edu/techdigest/content/techdigest/pdf/V35-N03/35-03-Howard.pdf.

Jung, Yoon. 2022. “Virtual Reality Simulation for Disaster Preparedness Training in Hospitals: Integrated Review.” *Journal of Medical Internet Research* 24 (1): e30600. https://doi.org/10.2196/30600.

Khanal, Shishir, Uma Shankar Medasetti, Mustafa Mashal, Bruce Savage, and Rajiv Khadka. 2022. “Virtual and Augmented Reality in the Disaster Management Technology: A Literature Review of the Past 11 Years.” *Frontiers in Virtual Reality* 3: 843195. https://doi.org/10.3389/frvir.2022.843195.

Koutitas, George, Scott Smith, and Grayson Lawrence. 2021. “Performance Evaluation of AR/VR Training Technologies for EMS First Responders.” *Virtual Reality* 25 (1). https://doi.org/10.1007/s10055-020-00436-8.

Le Noury, Peter J., Remco C. Polman, Michael A. Maloney, and Adam D. Gorman. 2023. “XR Programmers Give Their Perspective on How XR Technology can be Effectively Utilised in High-Performance Sport.” *Sports Medicine - Open* 9. https://doi.org/10.1186/s40798-023-00593-5.

Lochmannová, Alena. 2025. “Exploring the Role of Virtual Reality in Preparing Emergency Responders for Mass Casualty Incidents.” *Israel Journal of Health Policy Research* 14 (1). https://doi.org/10.1186/s13584-025-00681-9.

Meta. 2025a. “Depth API Overview.” *Meta for Developers*. Updated March 20, 2025. Accessed April 6, 2026. https://developers.meta.com/horizon/documentation/unity/unity-depthapi-overview/.

Meta. 2025b. “Get Started with Occlusions.” *Meta for Developers*. Updated February 18, 2025. Accessed April 6, 2026. https://developers.meta.com/horizon/documentation/unity/unity-depthapi-occlusions-get-started/.

Meta. 2025c. “Occlusions Overview.” *Meta for Developers*. Accessed April 6, 2026. https://developers.meta.com/horizon/documentation/unity/unity-depthapi-occlusions/.

Meta. 2026. “Controller Input and Tracking Overview.” *Meta for Developers*. Accessed April 6, 2026. https://developers.meta.com/horizon/documentation/unity/unity-ovrinput/.

Misra, Shalini, Patrick Roberts, and Matthew Rhodes. 2020. "Information Overload, Stress, and Emergency Managerial Thinking." *International Journal of Disaster Risk Reduction* 51: 101762. https://doi.org/10.1016/j.ijdrr.2020.101762.

Oregui, Xabier, Anaida Fernández García, Igor García Olaizola, Verónica Ruiz, Izar Azpiroz, Blanca Larraga-García, and Álvaro Gutiérrez. 2024. “Augmented Reality Interface for Adverse-Visibility Conditions Validated by First Responders in Rescue Training Scenarios.” *Electronics* 13 (18). https://doi.org/10.3390/electronics13183739.

Rydvanskiy, Ruslan, and Nick Hedley. 2021. “Mixed Reality Flood Visualizations: Reflections on Development and Usability of Current Systems.” ISPRS *International Journal of Geo-Information* 10 (2). https://doi.org/10.3390/ijgi10020082.

Saifi, Sanjay, and RAAJ Ramsankaran. 2024. “Augmented-Reality-Based Snow Visibility Simulation for Disaster Preparedness in the Western Himalayas.” *Proceedings of the International Association of Hydrological Sciences* 387. https://doi.org/10.5194/piahs-387-73-2024.

Sermet, Yusuf, and Ibrahim Demir. 2018. "Flood Action VR: A Virtual Reality Framework for Disaster Awareness and Emergency Response Training." Paper presented at the 2018 International Conference on Modeling, Simulation and Visualization Methods (MSV'18), Las Vegas, NV, August 2018. https://doi.org/10.1145/3306214.3338550.

Toner, Eric S. 2010. “Creating Situational Awareness: A Systems Approach.” *Medical Surge Capacity: Workshop Summary*. https://doi.org/10.17226/12804.

TRACLabs. 2018. “A Mixed Reality Training and Testing Facility for First Responders.” NIST. Accessed April 6, 2026. https://www.nist.gov/ctl/pscr/funding-opportunities/past-funding-opportunities/psiap-user-interface/mixed-reality.

Tsujimoto, Ryoma, Tomohiro Fukuda, and Nobuyoshi Yabuki. 2024. “Server-Enabled Mixed Reality for Flood Risk Communication: On-Site Visualization with Digital Twins and Multi-Client Support.” *Environmental Modelling & Software* 177: 106054. https://doi.org/10.1016/j.envsoft.2024.106054.

Uhl, Jakob Carl, Olivia Zechner, Anke Baetzner, Tanja Birrenbach, Sebastian Egger-Lampl, Helmut Schrom-Feiertag, and Manfred Tscheligi. 2025. "Mixed Reality Training for Medical First Responders: System Evaluation and Recommendations." *Virtual Reality* 29: 69. https://doi.org/10.1007/s10055-025-01144-x.

Unity Technologies. 2025a. “Set Up UI Canvases for XR.” XR Interaction Toolkit Documentation. Accessed April 6, 2026. https://docs.unity3d.com/Packages/com.unity.xr.interaction.toolkit@3.3/manual/ui-setup.html.

Unity Technologies. 2025b. “Tracked Device Graphic Raycaster Component.” XR Interaction Toolkit Documentation. Accessed April 6, 2026. https://docs.unity3d.com/Packages/com.unity.xr.interaction.toolkit@3.0/manual/tracked-device-graphic-raycaster.html.

Xi, Nannan, Juan Chen, Filipe Gama, Henry Korkeila, and Juho Hamari. 2024. “Virtual Experiences, Real Memories? A Study on Information Recall and Recognition in the Metaverse.” *Information Systems Frontiers* 27. https://doi.org/10.1007/s10796-024-10500-2.

Xu, Zhen, Yajun Yang, Yian Zhu, and Jingjing Fan. 2023. “Mixed Reality Drills of Indoor Earthquake Safety Considering Seismic Damage of Nonstructural Components.” *Scientific Reports* 13. https://doi.org/10.1038/s41598-023-43533-9.

Zhu, Yiqing, and Nan Li. 2021. “Virtual and Augmented Reality Technologies for Emergency Management in the Built Environments: A State-of-the-Art Review.” *Journal of Safety Science and Resilience* 2 (1). https://doi.org/10.1016/j.jnlssr.2020.11.004.