Development and Testing of Extended Reality Input Modalities for a Virtual Learning/Training Task

Stephanie G. Fussell, PhD, Benjamin J. Kwasa, PhD
Kent State University
Kent, OH
SFussel2@kent.edu, BKwasa@kent.edu

Quintin Oliver, Tyler Frost
Air Force Research Laboratory
Kettering, OH
Quintin.Oliver@us.af.mil, tyler.frost@baesystems.us

ABSTRACT

Immersive simulation technologies are integral to U.S. military training, with a focus on augmented reality (AR), mixed reality (MR), and virtual reality (VR)—collectively known as extended reality (XR). These technologies are explored for Distributed Mission Operations (DMO) and the Joint Simulation Environment (JSE). However, determining the most effective input modality for training tasks in dynamic virtual environments (VEs) remains a significant challenge, with existing studies often lacking comprehensive assessments and omitting MR or non-virtual (control) conditions.

To address this gap, the Air Force Research Lab in Dayton, OH, is developing a VE to evaluate various interaction methods for diverse training scenarios. This ongoing research evaluates the effectiveness of several input modalities for procedural training tasks. This paper details the development process of the VE, covering the modeling of flight instrument panels, the creation and integration of the input modalities, and the usability and preference results of the pilot test to complete a procedural training task. The results of the mixed-methods, within-subjects experiment indicate that dwell selection methods are the least preferred method, and that clicking with the trigger was most preferred regardless of the pointing method. However, the usability, sickness, and perceived workload scores were excellent, indicating that performance and preference may not correlate.

The research aims to provide practical guidance on XR device utilization in dynamic training scenarios, emphasizing the balance between physical interaction, scenario fidelity, and force readiness enhancement. Challenges encountered during the project are discussed, along with recommendations for creating similar VEs to improve understanding of optimal interaction methods in complex training environments.

ABOUT THE AUTHORS

Stephanie G. Fussell, Ph.D., is an Assistant Professor and Aeronautics Programs Coordinator at Kent State University's College of Aeronautics and Engineering. Her research investigates XR for aviation and aerospace training with a focus on usability, user experience, and transfer of training. Three times, she has been awarded an Air Force Research Laboratory Summer Faculty Fellowship with the 711th Human Performance Wing, Airman Systems Directorate, Warfighter Interactions & Readiness Division (SF.15.12.B0913). Dr. Fussell earned degrees in Aeronautics (BS, MS) and Aviation (PhD) at Embry-Riddle Aeronautical University.

Quintin Oliver, AFRL, is a Computer Scientist in the GRILL®. His work utilizes virtual, augmented, and mixed reality technologies to create rapid prototypes of environments focused on personalized training. In these environments, he leverages his interests of 3D modeling and artificial intelligence to create unique experiences.

Tyler Frost is a software engineer through BAE Systems at the Air Force Research Lab's (AFRL) Gaming Research Integration for Learning Lab (GRILL®). He graduated from Wright State University with a degree in computer engineering. His experience varies utilizing different areas of engineering such as development using game engines, and virtual/mixed reality systems integration.

Benjamin. J. Kwasa, Ph.D., is an Assistant Professor of Aerospace Engineering in the College of Aeronautics and Engineering at Kent State University. His research includes the investigation of XR as complex decision-making systems for both design and training. In addition, he conducts research in complex system design and heuristic optimization. Dr. Kwasa earned his bachelor's (Aerospace Engineering), master's (Systems Engineering), and doctorate (Aerospace Engineering) from Iowa State University.

Development and Testing of Extended Reality Input Modalities for a Virtual Learning/Training Task

Stephanie G. Fussell, PhD, Benjamin Kwasa, PhD
Kent State University
Kent, OH
SFussel2@kent.edu, BKwasa@kent.edu

Quintin Oliver, Tyler Frost
Air Force Research Laboratory
Kettering, OH
Quintin.Oliver@us.af.mil, tyler.frost@baesystems.us

INTRODUCTION

Immersive simulation technologies have long been used to train those in the U.S. military, such as part-task trainers, flight simulators that emulate the live-task environment, and high-fidelity Cave Automatic Virtual Environments (CAVEs). These technologies are ever-evolving, and several military branches (including the U.S. Army, Air Force, and Navy) are exploring the efficacy and usability of extended reality (XR) technologies for training. XR ranges in levels of immersion and realism, encompassing augmented reality (AR), mixed reality (MR), and virtual reality (VR). A review of the literature confirms that XR technologies offer learning and training benefits and that learning in virtual environments (VE) can be just as effective as training in traditional training environments due in part to the creation of highly realistic environments, providing interactive experiences that facilitate knowledge acquisition and information processing (Hamilton et al., 2021; Kaplan et al., 2021). Several studies found that training with XR technologies can positively impact engagement, presence, and motivation, positively impacting performance (Li et al., 2023; Makarova et al., 2023). These general findings are tempered by cases where the transfer of skills between virtual and live-task environments results in increased error rates, increased cognitive load, and decreased performance, making it clear that the design and development of XR technologies for learning and training must be carefully approached to promote knowledge and skill acquisition as well as an application outside of the VE (Petersen, et al., 2022; Rickel, 2023).

As the U.S. Air Force (USAF) expands Distributed Mission Operations (DMO) training to integrate Live, Virtual, and Constructive (LVC) concepts and initiatives to train the Warfighters, there is an opportunity to explore the training affordances that come with incorporating XR technologies into virtual simulators, gaming environments, and constructive models, as well as linking these applications with relevant live operational systems. The USAF can expand DMO LVC training and rehearsal to develop, test, and implement XR technologies into multiple applications. However, the state of the present technology, in terms of hardware and software, offers a variety of ways to interact with the VE. For example, higher-end XR products (including the Varjo XR3 and XTAL 3 MR) offer eye-gaze tracking, head tracking, and hand tracking in addition to compatibility with controller inputs. With so many choices comes flexibility, but also a question of the appropriateness of a given input modality for different interactions and VEs. Interactions that are unintuitive, unnatural for the task, or difficult to execute can cause frustration and negatively impact the training experience. The USAF and others can explore the design and development of XR technologies to enhance DMO LVC training and rehearsal concepts for Warfighters to further distribute training while leveraging cutting-edge technologies that will positively impact the training experience.

Studies that describe the advantages and drawbacks of learning/training in VLEs with XR technologies do not always consider the ramifications of how the student interacts with the environment in terms of input modality (also called input or object selection and selection technique). Although several studies compare user performance with different input modalities, they are not typically situated in a learning or training environment (Hou & Chen, 2021; Xu et al., 2022; Yu et al., 2018). Alternatively, studies that investigate input modalities in a learning/training environment often focus on the performance outcomes of a specific input modality as opposed to comparisons of multiple modalities to determine the best option for the learning/training task (Challenor et al., 2023; Ragan et al., 2015). In general, input modalities that feature controller-based selection mechanisms have the best results in terms of performance and user preference when compared to dwell selection mechanisms. Selection mechanisms that feature hand tracking have mixed results in terms of performance and user preference, most likely due to hardware and software limitations. The dearth of research comparing input modalities in a learning/training environment is evidence of a gap in the research that will be of interest to those who must ensure that the interaction between the student and the VE results in the positive transfer of knowledge, skills, and abilities (i.e., performance outcomes) between the VE and the real world.

Research Background and Goals

Dr. Fussell worked with a team at the Air Force Research Lab (AFRL) Gaming Research Integration for Learning Laboratory® (GRILL) in Kettering, Ohio, during the summers of 2022, 2023, and 2024. The GRILL is a Department of Defense (DoD) facility developing the next generation of training and our future workforce through STEM outreach in primary and secondary schools as well as universities across the country. As a part of the AFRL, the GRILL researches, evaluates, and leverages existing commercial off-the-shelf (COTS) hardware and software technologies to develop rapid prototypes, testbeds, and VEs to enable personalized training applications and research to improve the capabilities and decrease the time to readiness for our airmen and beyond. As the pace of technology changes, the requirement for adaptation increases as well, adaptation is most successful when one is armed with knowledge of these ever-growing systems of systems. The GRILL works with industry to create and maintain a knowledge base of existing and future game-based technologies, best practices, and any other pertinent information to STEM to share with the DoD and academia. This knowledge is shared via technical evaluations.

In the summer of 2022, the team developed and pilot tested a head-mounted display (HMD) comparison methodology (Fussell, 2023; Fussell, 2024). The methodology considers technical specifications, user design principles, desired KSAs (knowledge, skills, and abilities), learning outcomes, and actual user experience. The methodology is both replicable and adaptable to changing technologies and training goals that are also generalizable to similar use cases. The comparison methodology was expanded in the summer of 2023 to focus on input modalities and different levels of immersion in VR and MR. Development of the VE continued through the summer of 2024. The initial phase of the project involved creating interactive flight deck panels that utilized VR controllers, along with conducting a pilot test of the VR program comparing five VR input modalities at Kent State University (KSU).

In late summer 2024, we developed MR input modalities and tested various aspects of the VE based on performance metrics such as task efficiency, effectiveness of input modalities in executing operations, error rates, human factors principles, user experience, and user preference for input modality types in completing procedural training tasks. A full study utilizing eight input modalities is planned for the fall of 2024. Two key features of this study differentiate it from similar studies. First, the final study compares seven VR input modalities and a MR condition to a non-virtual (control) condition. In this context, MR refers to the ability to interact with real-world (physical) and simulated (virtual) components while completely immersed in the VE. Although similar studies may compare several virtual input modalities, they did not utilize an MR or a non-virtual condition. Second, most studies that compare input modalities are not set in a learning or training environment. This represents a huge gap in the literature, as XR technologies are rapidly being introduced to curricula and training programs for conceptual, procedural, and dynamic training purposes. Understanding if and how input modalities can impact KSA acquisition, mastery, and transference to the real world is important to ensure that learning outcomes are met while students receive the best quality training available.

This report discusses the development of the VE, including the modeling and integration of the different input modalities, and a pilot study of the comparative analysis of the input modalities as they impact the user's ability to complete the training task. The results of the final study will not be covered in this paper.

Research Questions and Hypotheses

1. How do different types of input modalities and levels of immersion impact the user's ability to complete a procedural task in XR?
   a. No significant difference in ability to complete the task in terms of effectiveness (as measured by error rate and number of attempts) between levels of immersion.
   b. No significant difference in ability to complete the task in terms of effectiveness (as measured by error rate and number of attempts) among input modality types.
   c. No significant difference in ability to complete the task in terms of efficiency (as measured as the time to complete a step in a procedural task, the time between steps, and the time to complete the procedural task) between levels of immersion.
   d. Significant difference in ability to complete the task in terms of efficiency (as measured as the time to complete a step in a procedural task, the time between steps, and the time to complete the procedural task) among input modality types.

2. How do types of input modalities and different levels of immersion impact the user's experience of executing a procedural task in XR?
   a. Significant difference in user experience of executing a procedural task (as measured by the E-SUS, NASA TLX) between levels of immersion.
   b. Significant difference in user experience of executing a procedural task (as measured by the E-SUS, NASA TLX) among input modality types.
3. How do different levels of immersion impact user cybersickness?
   a. No significant difference in cybersickness (as measured by the SSQ) between levels of immersion.

VIRTUAL ENVIRONMENT FRAMEWORK AND DEVELOPEMNT

The VR program being tested is a virtual checklist on flight deck panels made using Unreal Engine. Game engines such as Unreal Engine are a low-cost, oftentimes open-source, tool with large community support that can be leveraged to create VEs. The program includes a guided procedure to execute a series of tasks using different input modalities, including interacting with toggles, switches, and a selector knob that replicate a COTS flight switch panel. The VR program tests specific interactions with the VE. The guided procedural task was designed to reduce complexity, promote the successful execution of the steps, require no background knowledge of the procedure, mimic the actions taken in the live-task environment, and ensure navigation efficiency and interaction continuity across input modality conditions as much as possible. The VE is situated in an office space as opposed to a cockpit to ensure the participant is not distracted by the environment. The VE developed by the GRILL served as the medium to compare the input modalities. A focus of the program was that it included a task system to ensure that each participant was given a similar set of instructions, the actions taken by the user could be evaluated as correct or incorrect, and the measurements of the effectiveness of the input modality concerning the time each participant requires to complete the instructions for a given input modality could be taken from the program as output.

Task Engine

The task engine framework is a hierarchy of sessions, scenarios, tasks, and operations.

- Each trial of the VE is denoted as a "session" during which a user will utilize the different input modalities, one in each of the session's scenarios. Each session has an overall time limit to ensure that the maximum time for each participant is regulated. Upon expiring, the session timer ends the user's session and records all data from within the session to an external file for analysis.
- The input modality along with a list of one or more tasks constitutes an individual "scenario." Each scenario is unique to that session to ensure that the user will not experience the same input modality twice within the session. However, scenarios are generated at the time of usage to remove the requirement for researchers to generate unique scenarios that only differ by means of the input modality.
- "Tasks" are actions that require one or more operations such as turning on an aircraft's engines or preparing for landing. Each task has a name, maximum duration, and one or more operations. Each task is a standalone element of the VE. This is to allow researchers to add and remove tasks without reprogramming the system as a whole. With this system in place, it is possible to have as many or as few tasks in a scenario to fit research requirements. When a task is started, the framework reads the initial state of all operations and sets each element of the panel accordingly. This removes the need for the researcher to manually set the state of the virtual panel in each scenario for each participant. The order of tasks varies with each scenario to ensure the study participant does not simply memorize an order of events.
- Finally, an "operation" is an individual action such as changing a knob to a specific position or clicking a switch. Operations dictate how the task framework advances. Upon completion of an operation or the expiration of an operation timer, the framework attempts to move to the next operation in the task. If no operations are remaining in the current task, the framework attempts to advance to the next task. Similarly, if there are no remaining tasks in the scenario, the framework will attempt to move on to the next scenario or exit the VE if there are no further scenarios. Each operation has a name, description, maximum duration, and initial and correct states. The description is displayed to the user as an instruction informing the user what action needs to be taken to successfully complete the operation. At the beginning of an operation, the state of the flight deck panel is set to the initial state denoted by the operation entry. Every time the user takes an action to alter the state of the virtual panel, the framework checks the new state of the panel to determine if it matches the correct state of the current operation. If the states match, the VE will advance to the next operation if one is available. Figure 1 shows the physical flight deck panel with operations grouped by the task they are associated with.

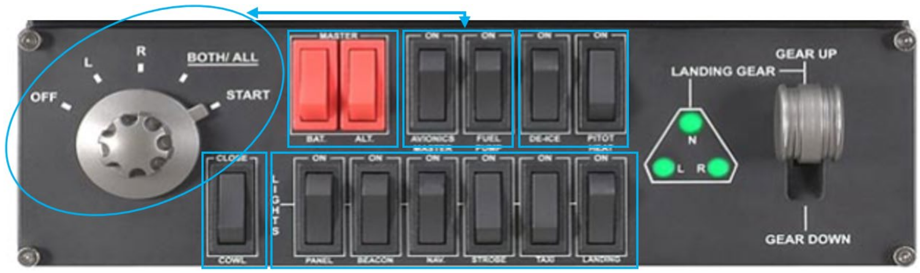

Figure 1. Physical Flight Switch Panel with Tasks Highlighted

Communication Between Tasks

The 3D virtual representation of the flight panel was sourced from the modeling and simulation community. The representation included a singular 3D model and an image file to use as a texture. The model was edited to separate each component such as the knob and switches. These new 3D models are what the users interact with in the VE. Each component behaves independently. When the framework system receives a user's simulated click, instruction is deferred to that component to handle the event if the action's target is one of the components; otherwise, no change occurs. Each component stores its state as well as the valid states to which it can transition. When a component successfully transitions to a new state, the framework first checks if that component is the target of the current operation, if so, the framework then checks if the new state is equal to the value set as the correct state as part of the operation. If both conditions are met, the system marks the action as correct. Otherwise, the action is marked as incorrect and will optionally revert the state of the component back to its previous state based on a setting selected by the researcher.

User Interface 3D Modeling and Visual Cueing

The user's view of the VE is a digital representation of the flight panel and a graphical user interface displaying information about the current scenario, task, and operation. The virtual panel emulates a COTS professional simulation flight switch panel. Throughout the session, the remaining time, as set by the researcher, is displayed for the current scenario and the current operation. Additionally, the current task number and total number of tasks are displayed in front of the user. This information increases the user's awareness of performance throughout the scenario. Directly above the virtual representation of the flight panel, the name and instructions for the current operation are displayed. These values are set by the researcher as part of the scenario development.

Throughout the session, various visual effects are optionally activated. Primarily, when the user is in a position to interact with one of the panel's components, an outline is drawn around that specific component to indicate that it is currently being hovered. By default, this appears as a yellow color. Additionally, if automatic hints are activated and a hint time is set, the yellow outline is replaced with a green color. The yellow hover visual cue always takes precedence over the green hint to ensure the user knows which component will be activated after the simulated click. The final visual cue is an optional label with the component's name that appears above the component when it is being hovered upon.

Scenario Generation and Tutorials

At the start of each session, three random input modalities are selected from the available options. Each modality chosen is unique to ensure that users do not receive the same modality twice within a single session. When a scenario starts, the user is presented with an instructional message describing the modality they will soon use. The participant is then instructed to don the HMD to continue in XR. Each modality has an associated tutorial that walks the user through the types of interactions that will be required of them during the upcoming scenario. These tutorial levels are designed to provide users without experience in XR reality with some familiarity with the XR controls implemented in the VE. Each tutorial is structured to ensure that the participant can become familiar with the controls without exposing them to the virtual panel before their first scenario begins. Once the tutorial is complete, the participants enter a new VE that presents them with the task framework and the virtual representation of the flight panel. At the end of the scenario, the user is then instructed to doff the HMD to provide a time for the researcher to include any intermediary steps that are external to the VE (i.e., completing a questionnaire and ensuring the user is ready and able to continue the study). Since each scenario can be executed regardless of the order or number of scenarios, a tutorial is always provided to the participant.

Logging

When a new session is started, the system requests a participant ID. This ID is used to create a log file specific to that user. A log entry is created when an advancement occurs within the VE to the next operation, task, or scenario and when any timer expires. Additionally, all user interactions with the virtual representation of the flight panel are recorded as correct and incorrect actions. When any of these events happen, the local time, UTC time, and seconds since the session started are recorded. Upon ending any scenario, all log entries are compiled into a comma-separated values file to be used for analysis later.

Input Modalities

The term input modality refers to how the user interacts with the VE in terms of the pointing method and the subsequent selection mechanism. Other terms include input selection, interaction method, object selection, and selection technique. The pointing method describes what is moved or manipulated in the VE and is usually one of four options: the head, the hand(s), the eyes, or the controller. Head, hands, and eyes are tracked using internal (i.e., HMD-based) and external (e.g., base stations) sensors. The selection mechanism denotes how the user confirms that an interaction should commence in the VE, once the pointing method has occurred. The three primary selection mechanisms in XR HMDs are dwell, click, and hand gesture, described in Table 1. As can be inferred, there are multiple combinations of pointing methods and selection mechanisms. Eight of these combinations have been identified for testing, detailed in Table 1. The italicized input modalities denote the five used in the pilot study and described in this report; the others will be used in the full study with an XR HMD.

Table 1. Description of Input Modalities in the Present Study and Literature

Input Modality	Description of Manipulation and Selection
Controller with button click ("controller + click")	Move the controller toward the virtual item to interact with, wait for the haptic or selection cue, and depress a finger on a controller button to interact. The button must be depressed and released for selection/interaction to occur.
Eye gaze with controller button click	The eye is tracked using HMD internal sensors. Move the eye over the virtual item to interact with, wait for the selection cue to appear, and depress a finger on the controller to select. The eye gaze may need to remain on the item for a set duration to trigger the selection cue.
Eye gaze with dwell	The eye is tracked using HMD internal sensors. Move the eye over the virtual item to interact with, wait for the selection cue to appear, and keep the gaze on the item for a set duration to confirm selection and appropriate in-program response.
Hand tracking with gesture in MR	Movement of the physical hand, mapped into the VE as a virtual hand, to the item to interact with and manipulate the item. No controllers are used, only physical hand movements, but the VE may have specific gestures for the selection/interaction to occur, such as a pinch (index finger-thumb movement) or grab (whole hand) gesture.
Head tracking with controller button click ("head tracking + click")	A visual cue, such as a crosshair, follows the movement of the head. Tilt/rotate the head to align the crosshair with the item to interact with, wait for the selection cue to appear, and depress a finger on the controller to select. The button must be depressed and released for selection/interaction to occur.
Head tracking with dwell ("head tracking + dwell")	A visual cue, such as a crosshair, follows the movement of the head. Tilt/rotate the head to align the crosshair with the item to interact with, wait for the selection cue to appear, and keep the crosshair on the item for a set duration to confirm selection and appropriate in-program response.
Laser pointer with button click ("laser + click")	Also called "ray casting," a laser/beam/ray emits from the controller. Point the laser at the item to interact with, wait for the haptic or selection cue, and depress a finger on the controller to select. The button must be depressed and released for selection/interaction to occur.
Laser pointer with dwell ("laser + dwell")	A laser/beam/ray emits from the controller. Point the laser at the item to interact with, wait for the selection cue to appear, then keep the laser on the item for a set duration to confirm selection and appropriate in-program response.

RESEARCH DESIGN AND METHODOLOGY

The study is a mixed-methods, within-subjects experiment. The pilot study occurred over two weeks. The participants were split into two groups based on which week they completed the study (i.e., Group A in week 1, Group B in week 2). Each participant completed a procedural training task using a Vive Pro 2 HMD. The independent variable was the type of input modality. The pilot study only tested 5, VR-specific input modalities (i.e., controller with button click, laser pointer with dwell or button click, head tracking with dwell or button click). The dependent variables of the pilot study are perceived workload, usability, and symptoms of cybersickness; issues for the development team to resolve were also collected. The convenience sample was 12 students enrolled in aeronautics and engineering programs at KSU and represents different ages, sexes, VR experience, educational background, work experience, etc. The input modalities were randomly assigned to each participant without replacement. Due to the small sample size of each group, the distribution of testing was not equal, and statistical analysis between groups could not occur. The larger sample size of the full study will allow for a more equal distribution. The full study will be conducted in the fall of 2024. It will include the time to complete the procedural task and error rate as dependent variables. A minimum sample size of 24 is needed to meet the following statistical analysis for a repeated-measures ANOVA for an effect size of 0.25, an alpha of 0.05, and a power of 0.8.

Apparatus and Materials

The VR program being tested was the VE described previously which included a set of tasks (checklist) to complete on a virtual flight deck panel. Figure 2 shows the VE with the flight deck panel, operation instruction, scenario and operation timing, and task sequencing.

Figure 2. Virtual Flight Switch Panel with Instructional, Timing, and Sequencing Information

The VR HMD used for the pilot study was the Vive Pro 2. The Varjo XR 3 or 4 will be used for the full study as it offers different levels of immersion (i.e., VR and MR). The input modalities are those described in Table 1. Valve Index controllers were used for the pilot study and will be used in the full study.

Measures and Instrumentation

In this study, "immersion" is a binary condition of MR and VR to represent full immersion (VR) or structured interaction with real-world objects and virtual hands (MR). Research question 1 will be determined using efficiency and effectiveness. Efficiency was measured as the time to complete a step in a procedural task, the time between steps, and the time to complete the procedural task. Effectiveness was measured by error rate and number of attempts to accomplish a step. Efficiency and effectiveness were logged within the program based on participant activity. This data was not collected during the pilot study, although the logging mechanism was tested, and changes were made based on researcher feedback for the full study. Research questions 2 and 3 are answered using questionnaires. The Extended System Usability Scale (E-SUS) assessed the overall usability of the technology using Likert scale items and open-ended responses. The NASA-TLX assessed workload along the subscales of mental demand, physical demand, temporal demand, performance, effort, and frustration. Participants were also asked which input modality was most and least preferred. Sickness was measured using the Simulator Sickness Questionnaire (SSQ) along the subscales of oculomotor, disorientation, and nausea. The questionnaires have been used in numerous usability studies, have demonstrated high construct and content validity as well as high face validity through iterative use, and have high reliability through use in different studies at different times and for different usability purposes. The pilot study focused on research question 2.

Procedures

The pilot study was broken into two phases. The goal for both groups was not only to test the usability of the VE but also to ensure that the questionnaires and procedures could meet the research questions. The scenarios included a tutorial on how to use the input modality to interact with a ball before interacting with the flight deck panel, as described in the section on VE development.

In part one (Group A), the participant completed three scenarios with randomized input modalities (no repetition, in sequence) while the researcher made notes on issues for the development team. In this phase, the appearance of hover text, visual cues, and dwell duration were manipulated so the participants could provide specific usability feedback. The procedure for Group A was: receive a briefing; complete the demographics form; complete three scenarios in VR; complete a questionnaire with items related to usability, experience, perceived workload, and preference; and receive a debrief. Based on participant feedback, some changes were made to the questionnaire (e.g., item clarity) and the VE (e.g., preferred dwell time, sequencing of the tutorial, verbiage of on-screen directions). The duration ranged from 25 to 50 minutes for each participant. The finalized questionnaires and scripted protocol were used for Group B.

The second phase emulates the full study procedures. The procedure for Group B was: receive a briefing, complete the demographics form, complete scenario 1, complete the questionnaire, complete scenario 2, complete the questionnaire, complete scenario 3, complete the questionnaire, and receive a debrief. The input modalities for each scenario were randomized without repetition. The questionnaires were the same for each scenario, with three items related to preference answered after the final scenario was completed. During this phase, a critical bug was discovered that caused scenarios to skip tasks. This issue was flagged but not fixed before the pilot study concluded. The duration ranged from 30 to 90 minutes for each participant.

RESULTS

The pilot study consisted of 12 students (2 female) ranging in age from 18 to 25 ($M = 21.67$, $SD = 1.37$). Eleven were undergraduate students, two were engineering students, 10 held a student or private pilot certificate, and one was an aviation management student. Three participants were owners of a VR headset, but only one claimed to use VR somewhat frequently. Overall, the VR familiarity of the participants was minimal, with 10 indicating infrequent use or “never used.”

Group A ($n = 7$) completed the three scenarios in sequence; thus, their data reflect their experience with the VE overall. The E-SUS scores ranged from 87.5 to 97.5 with an average score of 90.71 (out of 100, where an accepted average score is 68), indicating an excellent overall usability score. The weighted SSQ subscales of nausea, oculomotor, and disorientation symptoms were 2.73, 3.25, and 0.00, respectively, with an overall total sickness score of 2.67. Group A did not complete the NASA TLX items. This group was asked about their general confidence and comfort using a seven-point scale (seven as high/positive anchor). The participants reported that they were very comfortable using VR technology ($M = 5.71$, $SD = 1.11$) and felt that they could maneuver in the VE without much issue ($M = 5.57$, $SD = 1.27$). The participants also felt strongly in their abilities to confidently interact with the VR flight deck panels ($M = 6.29$, $SD = 0.76$) and to complete a basic procedural training task in the VE ($M = 6.43$, $SD = 0.79$). Finally, the participants felt strongly that the VR program allowed them to learn how to successfully interact with different panel components in the VE.

Group B ($n = 5$) completed each scenario separately followed by a questionnaire. All the input modalities were scored above average (68) in terms of usability, with all but the head tracking + dwell condition scored as “exceptional” (Table 2). Using the NASA TLX, the head tracking + click condition was rated “somewhat high” in terms of mental demand and frustration. The majority of the ratings were “medium” workloads across the input modality conditions and subscales, as seen in Table 3. Figure 3 shows that the head tracking + dwell and laser + click conditions had all SSQ scores of 0 (thus not represented on the graph), that only the conditions of controller + click and head tracking + click groups had minimal symptoms scores, and that the other scores are acceptably low.

Table 2. E-SUS Results

	Head tracking + Dwell (*n* = 4)	Head tracking + Click (*n* = 2)	Laser + Dwell (*n* = 4)	Laser + Click (*n* = 1)	Controller + Click (*n* = 4)
Score Mean	73.13	83.75	82.50	82.50	86.25
Score SD	7.47	1.77	15.94	n/a	14.79
Score Range	62.50 to 80	82.2 – 85	67.50 - 95	n/a	70 - 100

Table 3. NASA TLX Average Ratings and Overall Results

	Head tracking + Dwell (*n* = 4)	Head tracking + Click (*n* = 2)	Laser + Dwell (*n* = 4)	Laser + Click (*n* = 1)	Controller + Click (*n* = 4)
Mental Demand	16.25	35.00	9.00	15.00	13.00
Physical Demand	25.00	12.50	9.00	5.00	9.00
Temporal Demand	8.75	17.50	11.25	15.00	6.00
Performance	28.75	10.00	18.00	5.00	24.00
Effort	20.00	10.00	12.00	10.00	15.00
Frustration	10.00	32.50	9.00	10.00	11.00
Overall	20.08	21.83	12.87	10.00	18.42

Note. Green text indicates a low rating (0-9), blue text indicates a medium rating (10-29), and red text indicates a somewhat high rating (30-49). Other scores of the NASA TLX are high (50-79) and very high (80-100).

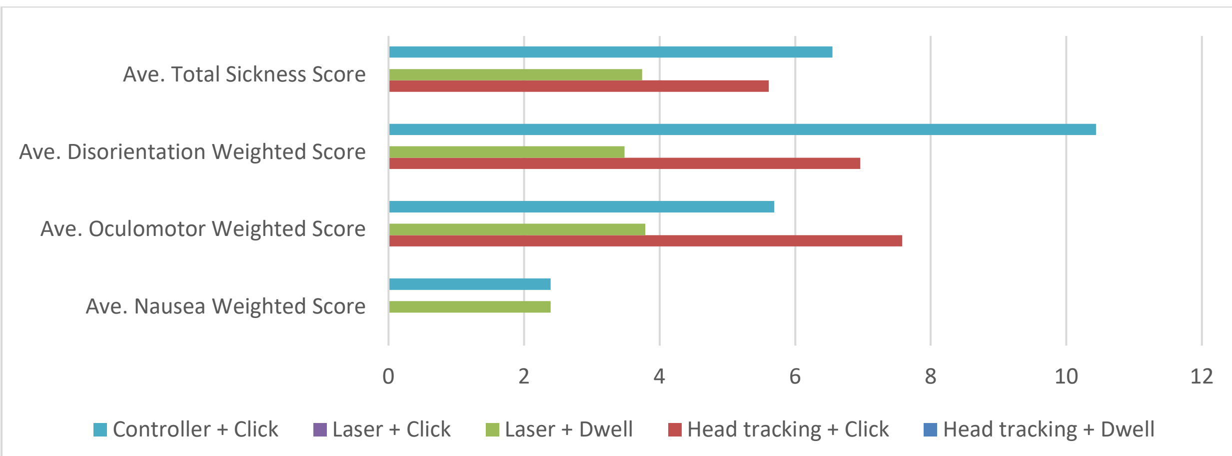

Figure 3. SSQ Results

Both groups were asked which input modality they preferred the most and the least. A summary of preferences is shown in Figure 4. The feedback can be categorized by point method (i.e., head tracking, laser, controller) and selection mechanism (i.e., dwell or click). Of the point methods, head tracking was the least favored option (*n* = 7) but controller + click and laser were favored equally (*n* = 4); the answer of one participant was not stated explicitly. For the selection method preferences, the modalities with dwell were the least preferred for both groups (*n* = 9) while a modality that included clicking the trigger was the most favored (*n* = 9). In general, the dwell selection mechanism was "least favorable because it felt as if I was not 'actually' in control of whatever I was doing or interacting with. Which from the previous question [referring to the most preferred], this type of input style felt as if it took away from the experience" (Participant 2809851). Additionally, the sentiment towards the input modalities with dwell was "It was odd not having to press any buttons. It was also the slowest of all 3" (Participant 1989606). Another participant said the head tracking + click modality was preferred the most because "It was the most efficient. It felt natural." In contrast, another participant said the head tracking + click modality was least preferred because "having to move my head back and forth felt like a little extra work" (Participant 1989606). The one participant who disliked the controller + click input modality cited calibration issues, which were discussed with the development team.

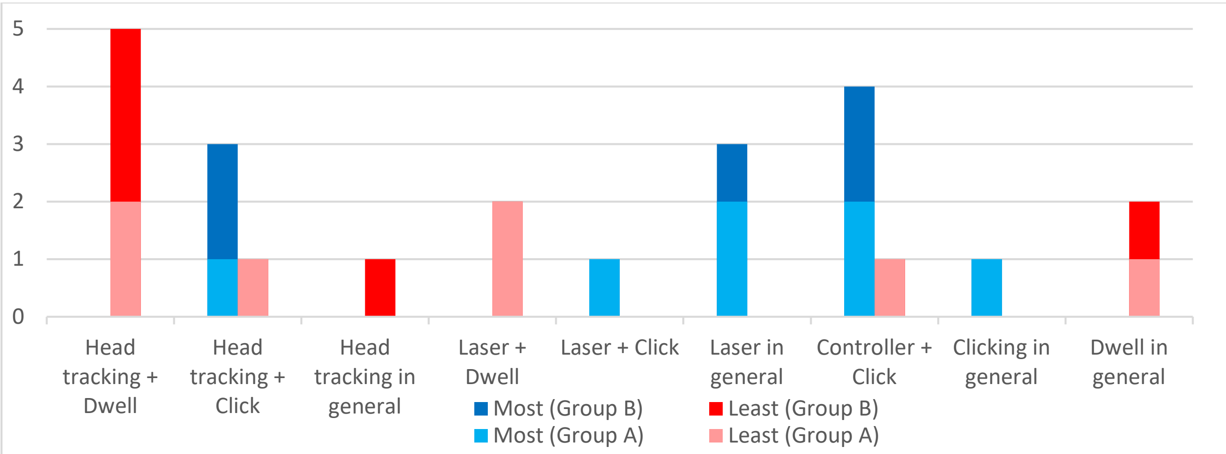

Figure 4. Summary of Input Modality Rated as Most and Least Preferred for each Group

The participants were also asked what they liked, disliked, and would modify about the program in general (Group A) and the different input modalities (Group B). Five participants from Group A commented positively on the usability of the VE, describing how easy it was to use, the clarity of directions, the simplicity of the interface, and that it was very beginner friendly. Participant 7889046 commented on the versatility, saying "I also like how you can input real scenarios and almost train for the real thing." Several Group A participants gave feedback to improve the VE, including comments on feedback, interactions, and the timer in the corner. The feedback for the head tracking point method received mixed reviews from Group B, as four participants said it was easy to learn how to use yet two stated that focusing the modality on a virtual object was also what they disliked about using the head tracking modality. Participant 7470012 stated they disliked "The need to point your face directly at what you want to interact with, and the delay in the interaction." The feedback for improvement included a need for more or better instructions on how to use the modality, which will be reviewed. Group B participants appreciated that the use of the laser included the movement of hands (as opposed to the head tracking) to interact with the VE, but the "dwell" component as the selection mechanism was not appreciated for the environment. Indeed, Participant 4083848 stated "I did not like the hover thing and I did not like the two second wait to flip a switch" and suggested, "Just having a button click." The feedback Group B gave after using the controller + click input modality reflected the desire to interact with the VE directly, with comments like "I liked the ability to click buttons in a VR world, and also that simplicity of it" (Participant 4083848). Yet, feedback on the controller button used for interaction was commented upon, such as "I would use a face button to interact (A) rather than the trigger" (Participant 7470012).

DISCUSSION

When developing task-based VEs, two of the most important factors to consider are the backend framework and the user interactions. The backend framework is structured to ensure that all operations are implemented in a standalone and state-agnostic fashion. However, the VE tracks its current overall state using interdependent logic sets that combine to create a unified representation of the flight panel. By deconstructing each task into distinct operations, each interaction with the VE becomes transactional enabling the return to previous states due to incomplete or incorrect actions. When including multiple input modalities, it is imperative that each modality performs the same set of actions in the VE. This was realized using a robust framework that forced every input modality to send either a "hover" or a "click" event to the VE as if the participant were using a computer mouse. For example, the dwell input modalities send a "hover" event to the VE when the participant places the emulated cursor on an object. When the emulated cursor remains in place for a preconfigured length of time, a "click" event is sent to the VE to trigger the interaction. These factors enable the separation of the VE into modules. In this case, the task framework, input modalities, and 3D models can be changed individually without affecting the other parts of the VE. This has allowed us the opportunity to expand the 3D model to include the multi-panel without impacting the current behavior of the VE.

Participants indicated that the most preferred input modalities had a click selection mechanism. Other studies also indicated that input modalities that feature controller-based selection mechanisms have the best results in terms of performance and user preference when compared to dwell selection mechanisms (Hellum et al., 2023; Hou & Chen, 202; Xu et al., 2022). The finding is reflected in participant feedback, such as "The most preferred type of input that I experienced was when I actually got to click the trigger and visually get a response whenever I hit one of the switches on the flight deck panel. I believe that this was my most favorable choice, because I could actually interact with whatever I was seeing (basically like real life where you can physically interact with the objects/panel)" (Participant 2809851). The usability of the VE proved to be quite high and is reflected in the confidence ratings of the participants to complete a task; the overall excellent scores of the E-SUS, SSQ, and NASA-TLX; and the open-ended responses. The positive feedback of both groups can be summarized in the words of Participant 5947088: "[The interface] was simple to use. The directions were clear. It reflected my exact motion of the hands and head movements." The lowest average usability score (head tracking + dwell) is still above the published average score and may reflect the overall sentiment that the dwell selection method is not an optimal choice for this environment. However, this finding is limited by the lack of efficiency data (i.e., timing and error rates). Group B's usability scores of the input modalities are also tempered by the small group sizes and unequal distribution, which must be addressed in the full study.

Reviewing the SSQ, both groups reported very low simulator sickness scores. The results of Group A indicate that, despite the back-to-back sequence of the scenarios, participants had negligible symptoms and a very low overall score. It is interesting to note that the highest scores of Group B were associated with the head tracking + click and controller + click input modalities. However, the unequal distribution does not allow for a deeper understanding. Further testing is needed to confirm if the higher weighted items are input modality or user-specific. Only the head tracking + click condition yielded somewhat high ratings of perceived workload in the subscales of mental demand and frustration. Based on participant feedback, the scores may be attributed to being uncertain of how to use input modality, which will be addressed in the tutorial of the full study. Of interest, the two input modalities with dwell did not receive higher scores, despite the clear preference disparity among the participants. This indicates that preference may not correlate with ability to complete the task. The efficiency and effectiveness scores from the full study will provide more information. Overall, the results are encouraging.

FUTURE RESEARCH

Beginning in the summer of 2023 through the spring of 2024, a VE (the flight deck in an office space, featuring different input modalities) was developed. Of the eight planned modalities, five were created and pilot-tested. In the summer of 2024, the development of the VE will continue to include a second flight deck panel and additional input modalities for a total of eight. The 2024 and 2025 experimental design includes a very simple procedural task (e.g., flipping switches, turning knobs) that is limited by the options in the 2 panels. The inability to draw correlations from the pilot study is a limitation that will be addressed in the full study. The next steps are:

1. Incorporate a second flight panel (multi-panel) and expand scenarios.
2. Incorporate the non-virtual condition and MR input modality.
3. Complete the full study and address all research questions.

The full study will occur during the fall of 2024. If funding through the Summer Faculty Fellowship is received for 2025, work will continue to expand the VE to create a realistic cockpit with more flight deck panels as opposed to an office space. The goal of the 2024 and 2025 projects is to understand which input modalities are best suited for procedural training in a complex environment (i.e., a cockpit). In the expanded VE, we will simulate a checklist or procedural trainer with more dynamic interactions, such as an engine start-up. The 2025 project will also compare the different input modalities to complete the procedure to answer the research questions proposed in this paper. The development team is confident that the full study will accurately answer the research questions and provide a greater understanding of which input modalities are best suited for a procedural training task.

ACKNOWLEDGEMENTS

The sponsor of this project was the Warfighter Interactions and Readiness Division of the U.S. Air Force (USAF). The author wishes to thank Dr. Lorraine Borghetti and Dr. Winston "Wink" Bennett for their support in this project, and the GRILL XR development team for their dedication to creating VEs to expand this research which supports not only Warfighter and military training but also civilian aviation training.

REFERENCES

Challenor, J., White, D., & Murphy, D. (2023). Hand-controlled user interfacing for head-mounted augmented reality learning environments. *Multimodal Technologies and Interaction, 7*(6), 55. https://doi.org/10.3390/mti7060055

Fussell, S. G. (2024). Review of VR technical specification requirements for a procedural training task. *Journal of Aviation/Aerospace Education & Research*, *33*(3). https://doi.org/10.58940/2329-258X.2064

Fussell, S. G. (2023). A review of user experience requirements for a procedural training task in virtual reality. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Washington, DC, Oct 23-27, 2023. https://journals.sagepub.com/doi/10.1177/21695067231192437

Hamilton, D., McKechnie, J., Edgerton, E. et al. (2021). Immersive virtual reality as a pedagogical tool in education: A systematic literature review of quantitative learning outcomes and experimental design. *Journal of Computers in Education, 8,* 1-32. https://doi.org/10.1007/s40692-020-00169-2

Hellum, O., Kersten-Oertel, M. & Xiao, Y. (2023). Assessment of user-interaction strategies for neurosurgical data navigation and annotation in virtual reality. *Virtual Reality 27*, 1345–1355. https://doi.org/10.1007/s10055-022-00740-5

Hou, W. & Chen, X. (2021). Comparison of eye-based and controller-based selection in virtual reality. *International Journal of Human–Computer Interaction, 37*(5), 484-495. https://doi.org/10.1080/10447318.2020.1826190

Kaplan, A. D., Cruit, J., Endsley, M., Beers, S. M., Sawyer, B. D., & Hancock, P. A. (2021). The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: A meta-analysis. *Human Factors, 63*(4), 706–726. https://doi.org/10.1177/0018720820904229

Li, Y., Ch'ng, E., & Cobb, S. (2023). Factors influencing engagement in hybrid virtual and augmented reality. *ACM Transactions on Computer-Human Interaction, 30*(4), 65:1-27. https://doi.org/10.1145/3589952

Makarova, I., Mustafina, J., Boyko,A., Fatikhova, L., Parsin ,G., Buyvol, P., & Shepelev, V. (2023). A virtual reality lab for automotive service specialists: A knowledge transfer system in the digital age. *Information, 14*(3), 163:1-26. https://doi.org/10.3390/info14030163

Petersen, G. B., Klingenberg, S., & Makransky, G. (2022). Pipetting in virtual reality can predict real-life pipetting performance. *Technology, Mind, and Behavior, 3*(3: Autumn). https://doi.org/10.1037/tmb0000076

Ragan, E. D., Bowman, D. A., Kopper, R., Stinson, C., Scerbo, S., & McMahan, R. P. (2015). Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task. In *IEEE Transactions on Visualization and Computer Graphics, 21*(7), 794-807. https://doi.org/10.1109/TVCG.2015.2403312

Rickel, E. (2023). *Effectiveness and user experience of augmented and mixed reality for procedural task training* (Publication No. 725) [Master's thesis, Embry-Riddle Aeronautical University]. Scholarly Commons. https://commons.erau.edu/edt/725/

Xu, W., Meng, X., Yu, K., Sarcar, S., & Liang, H. (2022). Evaluation of text selection techniques in virtual reality head-mounted displays. *Proceedings of the 2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR).* Singapore, Singapore, 2022 131-140. https://doi.org/10.1109/ISMAR55827.2022.00027

Yu, D., Liang, H., Lu, F., Nanjappan, V., Papangelis, K., & Wang, W. (2018). Target selection in head-mounted display virtual reality environments. *Journal of Universal Computer Science, 24*, 1217-1243. https://api.semanticscholar.org/CorpusID:53281261