Comparing Input Modalities in Extended Reality for a Virtual Learning/Training Task

Stephanie G. Fussell, Summer Rebensky, Samantha K. B. Perry
Aptima, Inc.
Fairborn, OH
Sfussell@aptima.com, srebensky@aptima.com, sperry@aptima.com

Quintin Oliver, Tyler Frost, Stephen McGee
US Air Force Research Laboratory - GRILL
Dayton, OH
quintin.oliver@us.af.mil, tyler.frost@baesystems.us, stephen.mcgee.3@us.af.mil

Benjamin J. Kwasa
Kent State University
Kent, OH
bkwasa@kent.edu

ABSTRACT

Military training in the U.S. increasingly utilizes immersive simulation technologies to enhance Distributed Mission Operations (DMO) and the Joint Simulation Environment (JSE). As extended reality technologies (XR)—including augmented reality (AR), mixed reality (MR), and virtual reality (VR)—evolve, their hardware and software capabilities are continuously explored. These technologies offer a variety of input modalities for interacting with and selecting objects in virtual environments. Despite the range of interaction methods available, research on the optimal input modality for training tasks in dynamic environments is insufficient.

The Air Force Research Lab's Gaming Research Integration for Learning Laboratory in Dayton, OH, has developed a testbed to evaluate the efficiency and effectiveness of seven input modalities to address this gap. The virtual environment simulates flight deck panels to serve as a procedural task trainer to collect data on timing, error rate and type, usability, and user preference. This study expanded the research of Fussell et al. (2024) to include two eye-tracking modalities and additional tasks using a second virtual flight panel. Seven input modalities were compared using a mixed-methods, within-subjects experimental design. Although ongoing, current results suggest that input modalities that rely on a dwell selection mechanism can cause frustration, as participants desire a more interactive approach. The eye gaze modalities have mixed results due to inaccurate tracking for some, causing frustration as well as oculomotor symptoms. The evidence suggests that eye tracking, especially if combined with a dwell condition, presents challenges that may render it ineffective in complex training environments. The research aims to understand how different input modalities affect a user's ability to complete procedural tasks in XR and how they influence user experience. The take-aways can be applied to the design and development of complex XR training simulators with an input modality that promotes correct behaviors with increased usability aspects.

ABOUT THE AUTHORS

Stephanie G. Fussell, PhD, is a Scientist with Aptima, working closely with the Air Force Research Laboratory's (AFRL) Gaming Research Integration for Learning Laboratory® (GRILL®). Her research focuses on XR training competency, efficiency, and effectiveness; meeting user experience requirements; and demonstrating transfer of training.

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 GRILL. His experience varies utilizing different areas of engineering such as development using game engines, and virtual/mixed reality systems integration.

Summer Rebensky, PhD, is a Senior Scientist at Aptima, Inc., leading AFRL work focused on training solutions for the modeling and simulation environment, managing a portfolio of work focused on evaluating simulation training capabilities and informing readiness through virtual training data. She leads efforts to develop, test, and implement modern training solutions to improve and measure human performance.

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. Samantha Perry, Division Director and Principal Scientist, Training & Learning Readiness Division, Aptima, Inc. has supported dozens of projects primarily associated with the AFRL and Army Research Institute (ARI), serving as principal investigator (PI) or project manager (PM) on many of those efforts. She oversees a multidisciplinary, multi-domain, multi-million-dollar portfolio of projects, integrating scientists, engineers, data analysts, and many others in providing solutions for intelligence analysts, training specialists, and operational personnel and leaders.

Stephen McGee, AFRL, is a Research Scientist and Team Lead of the GRILL, where he spearheads innovative research projects and advancements and leads the Just-in-Time Multi Mission Airmen/Warfighters (JITMMA/W) Program. His work focuses on integrating technology and innovation to ensure that warfighters receive the most effective and efficient training possible.

Comparing Input Modalities in Extended Reality for a Virtual Learning/Training Task

Stephanie G. Fussell, Summer Rebensky
Aptima, Inc.
Fairborn, OH
Sfussell@aptima.com, srebensky@aptima.com

Quintin Oliver, Tyler Frost, Stephen McGee
US Air Force Research Laboratory - GRILL
Dayton, OH
quintin.oliver@us.af.mil, tyler.frost@baesystems.us, stephen.mcgee.3@us.af.mil

Benjamin J. Kwasa
Kent State University
Kent, OH
bkwasa@kent.edu

INTRODUCTION

Immersive simulation technologies have been a staple in military training contexts for decades. In aviation especially, flight training simulators have evolved from the Link trainer of the 1930-40s to cutting-edge training devices equipped with motion platforms and head-mounted displays (HMDs) that offer high visual fidelity. The addition of extended reality technologies (XR)—which includes augmented reality (AR), virtual reality (VR), and mixed reality (MR)—into flight training programs has positively impacted training efficiency and effectiveness in both military and general aviation contexts (Kaplan et al., 2021; Urban & Pritchard, 2024). The U.S. Air Force (USAF) is advancing Live, Virtual, and Constructive (LVC) concepts and initiatives within Distributed Mission Operations (DMO), including integrating XR technologies into training. XR enables innovative training and rehearsal techniques and facilitates integration with live operational systems. However, the addition of XR into a complex training environment comes with the challenge of ensuring the design of the virtual elements—and how the user interacts with them—is efficient, evokes natural movement, does not increase cognitive load, and does not create incorrect (negative) behaviors and attitudes that will then transfer to the real-world (Rickel, 2023).

The continued integration of XR into training programs requires frequent assessment as technologies progress and emerge in terms of hardware and software capabilities. For example, current trends in flight training utilize XR technologies to provide an MR cockpit with both physical and virtual instrumentation as well as a simulated environment. The use of passthrough cameras allow the pilot to interact with the physical and virtual elements while wearing an HMD as opposed to looking at a projected image. The XR technology may also feature a variety of ways to interact with and select virtual elements through eye, head, hand, and controller tracking. The choices offer flexibility but also add complexity to design and develop a virtual environment (VE) for training that promotes the development and then transfer of knowledge, skills, abilities, and other characteristics (KSAOs) between the virtual and real-world environments. With the increase in ways to interact with an XR environment comes a research challenge to understand how input modalities (also called input or object selection and selection techniques) impact training efficiency and effectiveness. There is a gap in the literature that explores the advantages and challenges of XR training environments while also considering available input modalities (Fussell et al., 2024; see also Hou et al., 2022; Rebensky et al., 2023; Xu et al., 2022).

Research Background

To bridge this gap, a VE was developed as a testbed to explore the modeling, integration, and usability of different input modalities to complete a procedural task. The testbed allows researchers and decision makers to understand the benefits and challenges of eight input modalities, with the ability to compare efficiency and effectiveness measures against a purely physical environment (e.g., a control condition). The work of Fussell et al. (2024) presented the results of a pilot study of the testbed. The present report includes results from the first round of data collection, as more data is needed to meet the minimum sample size to conduct an ANOVA.

Input modalities refer to how a user interacts with a VE through a pointing method and the subsequent selection mechanism. The pointing method is what the user moves or manipulates within the VE—generally, the head, hand(s), eyes, or controller—and is tracked using internal or external sensors. The selection mechanism is how the user confirms their action and that an interaction can begin—generally, a click, dwell/hover, or hand gesture. Table 1 describes eight input modalities from the extent literature and utilized in the testbed, although the italicized input modality was not tested for the results presented in this report.

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 (1 sec in present study) to confirm selection and appropriate in-program response.
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 (1 sec in present study) 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 (1 sec in present study) 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.*

It is important to determine the effects of different input modalities on the acquisition, mastery, and real-world application of KSAOs, obtained within an XR training program, to guarantee that trainees achieve the intended learning outcomes and receive the highest quality training. The testbed allows researchers to compare the input modalities using performance metrics such as task efficiency, effectiveness of input modalities in executing operations, and error rates. These metrics are collected within the architecture of the program. User experience aspects of sickness, immersion, workload, and preference are also gathered using questionnaires.

Research Questions and Hypotheses

The following research questions and hypotheses will be tested using the XR input modality testbed. The current study considers research questions 2 and 3 only using VR conditions of the seven input modalities.

1. How do different types of input modalities impact the user's ability to complete a procedural task in XR?
   H1. 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.
   H2. 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) 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?
   H3. No 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 types of input modalities impact user cybersickness?
   H4. No significant difference in cybersickness (as measured by the SSQ) among input modality types.

TESTBED DEVELOPMENT

The design and development of the testbed began in 2023 by a team of engineers and researchers at the Air Force Research Lab's (AFRL) Gaming Research Integration for Learning Laboratory® (GRILL®) in Kettering, Ohio. The GRILL is a Department of Defense (DoD) facility focused on developing the next generation of training applications, and the future workforce, through STEM outreach to local primary and secondary schools and nationwide higher education institutions. The GRILL develops rapid prototypes, testbeds, and VEs by researching, evaluating, and leveraging government- and commercial-off-the-shelf (GOTS, COTS) hardware and software technologies. These VEs enable personalized training applications and promote research opportunities to improve the capabilities and decrease the time to readiness for our warfighters with takeaways appliable to the civilian side. The GRILL also works with government, industry, and academia to generate and maintain a knowledge base of current and future game-based technologies, best practices, and any information relevant to STEM education, shared via technical evaluations. The design and development, as well as history of the project, are discussed in more detail in reports by Fussell (2023, 2024) and Fussell et al. (2024), although an overview of the testbed framework is provided here.

The VE replicates two COTS flight deck panels made using Unreal Engine as a testbed to compare users' ability to complete a procedural task efficiently and effectively using a randomly assigned input modality. The flight deck panels are situated in a simple office space as opposed to a cockpit to reduce the chance of distraction. The testbed is a guided procedural trainer wherein the user executes a checklist with 15 tasks by interacting with toggles, switches, and selector knobs. As noted in Fussell et al. (2024), "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" (p. 4). The framework consists of sessions, scenarios, tasks, and operations. Figure 1 shows the VE with the flight deck panels, operation instruction, scenario and operation timing, and task sequencing.

- Each trial of the VE is denoted as a "session" during which a user will utilize three of the input modalities in a unique scenario with a list of tasks and operations.
- A "scenario" refers to the progression through the VE by completing tasks and operations for single input modality. At the beginning of each scenario the user is presented with a tutorial to familiarize them with the types of interactions that will be required of them during the upcoming scenario. Every input modality starts the user with a tutorial to ensure they have the familiarity needed to interact with the VE. After the tutorial, the environment changes to include the virtual flight panels and the tasks.
- "Tasks" are actions that require one or more operations using the virtual flight panels to emulate the actions required of a flight procedure or checklist. The program includes a task system to ensure that each participant is given a set of instructions with the same tasks but in varying order to ensure the study participant does not memorize an order of events.
- An "operation" is an individual action such as changing a knob to a specific position or flipping a switch, which may be grouped to form a task. Each operation is presented to the user as a text-based instruction.

Throughout the session, various visual effects are optionally activated. When the user hovers on a panel component, a yellow outline is drawn around that specific component to indicate that the user can interact with it. Automatic hints can be activated and if a hint timer is set, the element to interact with is outlined in green. The yellow hover visual cue always takes precedence over the green hint to ensure the user knows the element with which they will interact. A final visual cue is an optional label with the component's name appearing above the component when the user hovers on it. Figure 2 shows the VE with the yellow and green visual cues.

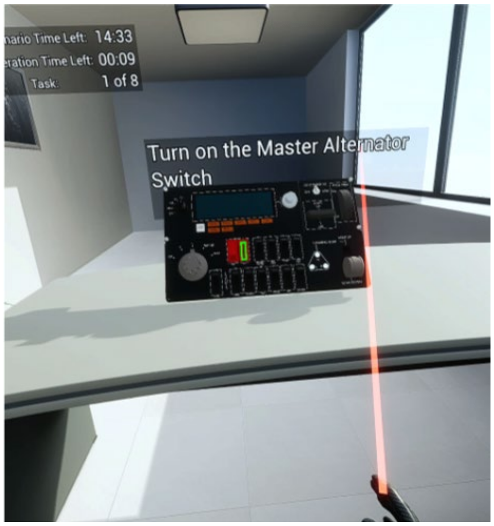

Figure 1. Virtual Environment with a Laser Modality Showing Operation Instruction, Timing, and Task Sequencing

Figure 2. Virtual Environment with a Headtracking Modality Showing the Yellow Visual Cue and the Green Hint Visual Cue

Once an operation is finished or its allotted time of 30 seconds runs out, the system will proceed to the next operation within the task. The system advances to the next task if there are no further operations left in the task at hand. When all tasks within the current scenario are completed, the system will transition to the next scenario, or it will end the session if there are no additional scenarios to undertake. The framework captures every interaction taken by the user and saves it to a log file to be analyzed later. The actions are evaluated as correct or incorrect. Timing is logged as a measurement of the effectiveness to compare how long each scenario, task, and operation takes among the input modalities. At the end of the session, the VE records all data from within the session to an external file for analysis.

RESEARCH DESIGN AND METHODOLOGY

The study uses a mixed-methods, within-subjects, experimental approach. The first round of data collection occurred at the Kent State University College of Aeronautics and Engineering (KSU CAE) March–April, 2025. The convenience sample included KSU CAE students, representing different ages, sexes, experience with XR, work and educational background, etc. Participants were recruited using Canvas Learning Management System announcements and emails. A minimum sample size of 43 is needed for statistical analysis using a repeated measure, within factors ANOVA for an effect size of 0.25, an alpha of 0.05, and a power of 0.8. The experimental design randomly assigned to each participant to three of the seven input modalities without replacement (see Table 1). A pseudo-random assignment approach will be considered to ensure an even distribution of assignment during the next round of data collection. The testing size for each input modality was not evenly distributed, limiting the types of analyses that can occur between groups. The independent variable was the type of input modality. The dependent variables of the study are preference, perceived workload, usability, and symptoms of cybersickness. Performance metrics, including the time to complete the procedural task and error rate, will be analyzed as dependent variables when the required sample size is reached.

Apparatus and Materials

Each participant completed a procedural training task using a Varjo XR-4 HMD with associated controllers. The VE testbed requires the participant to complete a procedural training task using two flight deck panels. The testbed was created in Unreal Engine 5.4 and was accessed using an Alienware gaming computer. Figure 3 shows a participant completing a tutorial (left) and interacting with the VE (right) using the Laser + Click input modality.

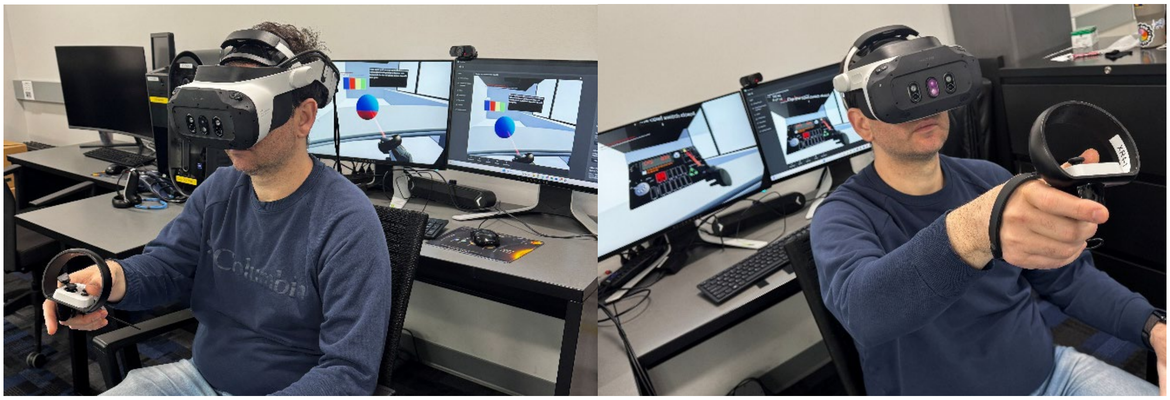

Figure 3. A participant completes a scenario. Photo shared with permission.

Measures and Instrumentation

Research questions 2 and 3 were addressed in the current report. Research question 2 was answered using questionnaires. The Extended System Usability Scale (E-SUS; Harper & Dorton, 2021; see also Brooke, 1995) was used to assess the overall usability of the technology using Likert scale items and open-ended responses. The NASA-TLX (Hart & Staveland, 1988) was used to measure workload along six subscales of mental demand, physical demand, temporal demand, performance, effort, and frustration. Participants were also asked which input modality they preferred the most and least. Research question 3 was answered using the Simulator Sickness Questionnaire (SSQ; Kennedy et al., 1993), which has subscale scoring of oculomotor, disorientation, and nausea as well as a total score. These questionnaires have been used in numerous studies, demonstrating high construct, content, and face validity as well as reliability through iterative use.

Procedures

The participants first received a briefing, reviewed an informed consent document, and completed a demographics form. The participants then donned the XR-4 HMD and completed the automatic eye calibration. Three input modalities scenarios were randomly chosen upon starting the VE. Each scenario included a tutorial on how to interact with the VE using the input modality before starting the procedural training element with the flight deck panels. When the scenario was completed, the participants removed the XR-4 HMD and completed the E-SUS, SSQ, and NASA TLX questionnaires. This process repeated three times, for each input modality. The participants also gave feedback on which input modality they preferred the most and least after completing the three scenarios. Finally, the participants received a debrief. The duration ranged from 30 to 70 minutes for each participant.

RESULTS

The study consisted of seven engineering students and six aeronautics students (*n* =13, 3 female) ranging in age from 19 to 36 (*M* = 23.85). Ten said they had a level of familiarity with using a VR HMD (five strongly so) while three participants said their familiarity was "neutral." One participant owned a VR HMD. Despite many having familiarity with VR, only four claimed that they used VR frequently; one participant used VR "sometimes," six did not use VR frequently, and one had never used VR. The participants were also asked if they were prone to motion sickness. Two responded with "sometimes" while four said "not frequently" and seven said "never."

Each participant completed a scenario then immediately responded to the E-SUS, SSQ, and NASA TLX questionnaires. The results for each questionnaire, by input modality, are presented in Tables 2, 3, and 4. Two of the input modalities, Controller + Click and Eye gaze + Dwell, show generally negative results (i.e., below-average usability, increased sickness symptoms, and higher workload) across the questionnaires. Only the Headtracking + Click input modality had generally positive results (i.e., excellent usability, negligible sickness symptoms, and lower workload) across the questionnaires.

Four of the input modalities had excellent usability scores, and one was above average (shown in green and blue, respectively, in Table 2). Eye gaze + Dwell had the highest weighted scores across all SSQ subscales, although Eye gaze + Click had no symptoms reported by the participants (shown in red and green, respectively, in Table 3). Six of the input modalities demonstrated high and somewhat high mental demand, effort, and overall workload scores (shown in red and orange, respectively, in Table 4). Only the Head tracking + Click modality had a medium overall rating and acceptable ratings across the six workload subscales. Controller + Click and Eye gaze + Dwell had the highest workload scores across the subscales and overall.

A review of the data in SPSS using Levene's test confirmed unequal samples sizes and variance among the input modalities sample sizes. ANOVAs using the Games-Howell posttest comparing the usability, sickness, and workload scores among input modalities were inconclusive, because the minimum sample size was not reached to perform an ANOVA with an effect size of 0.25, an alpha of 0.05, and a power of 0.8.

Table 2. E-SUS Results

	Score Mean	**Score SD**	**Score Range**
Controller + Click (*n* = 8)	66.88	24.01	32.5–100
Eye gaze + Click *(n* = 4)	88.13	6.25	80–95
Eye gaze + Dwell (*n* = 5)	52.00	33.19	12.50–97.50
Head tracking + Click (*n* = 4)	91.88	11.43	75–100
Head tracking + Dwell (*n* = 4)	88.13	6.25	22.50–90
Laser + Click (*n* = 8)	85.31	15.26	57.50–100
Laser + Dwell (*n* = 6)	76.67	18.28	50–100

Note. An average score is 68.00. Green text indicates an "excellent" score (85-100), blue text indicates a "good" score (71-84), orange text indicates an "ok" score (51-70). There are also scores of "poor" (25-50) and "worst imaginable" (0-24). See Bangor et al. (2009) and Sauro (2018).

Table 3. SSQ Results

	Ave. Nausea Weighted Score	**Ave. Oculomotor Weighted Score**	**Ave. Disorientation Weighted Score**	**Ave. Total Sickness Score**
Controller + Click (*n* = 8)	3.58	11.37	15.66	11.22
Eye gaze + Click *(n* = 4)	0.00	0.00	0.00	0.00
Eye gaze + Dwell (*n* = 5)	20.99	34.87	38.98	35.90
Head tracking + Click (*n* = 4)	4.77	3.79	0.00	3.74
Head tracking + Dwell (*n* = 4)	7.16	11.37	10.44	11.22
Laser + Click (*n* = 8)	2.39	7.58	5.22	6.08
Laser + Dwell (*n* = 6)	6.36	10.11	2.32	8.10

Note. Green text indicates negligible or no symptoms (0-4), blue text indicates minimal symptoms (5-10), orange text indicates significant symptoms (11-15), and red text indicates that symptoms are a concern (16-20) with score above 21 indicating an issue with the simulator. The categories are for total scores. See Kennedy et al. (2003).

Table 4. NASA TLX Average Ratings and Overall Results

	Mental Demand	Physical Demand	Temporal Demand	Performance	Effort	Frustration	Overall
Controller + Click (*n* = 8)	50.00	43.13	32.50	50.63	50.00	42.50	52.08
Eye gaze + Click (*n* = 4)	58.75	25.00	22.50	25.00	40.00	40.00	43.58
Eye gaze + Dwell (*n* = 5)	65.83	42.50	48.33	39.17	65.00	51.67	58.94
Head tracking + Click (*n* = 4)	29.00	14.00	11.00	21.00	14.00	7.00	19.93
Head tracking + Dwell (*n* = 4)	46.25	35.00	15.00	37.50	33.75	41.25	37.92
Laser + Click (*n* = 8)	30.00	31.88	17.50	27.50	37.50	18.75	31.13
Laser + Dwell (*n* = 6)	34.29	39.29	13.57	21.43	42.86	23.57	34.29

Note. Green text indicates a low rating (0-9), blue text indicates a medium rating (10-29), orange text indicates a somewhat high rating (30-49), red text indicates a high rating (50-79). The final rating option is very high (80-100).

The participants were asked to identify which input modality they preferred the most and the least using open-ended feedback, summarized in Table 5. The top two most preferred input modalities were Laser + Click (*n* = 5) and Head tracking + Click (*n* = 4). The former was noted by one participant as easy to aim and quick to confirm the interactions (using the controller trigger), while the latter was described as offering natural control via head movement that was also easy to keep in a single spot and aid interaction with the virtual objects. The two least preferred input methods were Eye gaze + Dwell (*n* = 5) and Laser + Dwell (*n* = 4). The Eye gaze + Dwell input modality was noted as inconsistent, difficult to ensure the tracking area stayed for the entire dwell time (1 second), and inaccurate. The Laser + Dwell input modality was noted as slow by a few participants, although this feedback is more concerned with the selection mechanism rather than an attribute of the laser pointer.

The feedback can also be reviewed by point method (i.e., head tracking, eye tracking, laser, or controller) and selection mechanism (i.e., dwell or click). The laser point methods were reported as the most preferred as well as the least preferred overall. The open-ended feedback indicates that while the laser offered easy accuracy, the design of the VE could hinder interaction—as in, the desk blocked the laser and the panels were situated high up in the VE, causing ergonomic challenges. The eye gaze point methods were reported as challenging and inaccurate, and some participants also noted that it caused adverse physical effects. Incorrect eye calibration caused further challenge for the eye tracking point method: some participants required a refreshed calibration after the tasks commenced while others physically moved their chair to be closer to the virtual elements. In general, input modalities with the click selection mechanisms were the most preferred (*n* = 12) while the dwell selection mechanism was the least preferred (*n* = 10). Interacting with the controllers allowed for quick and easy selection confirmation, with one participant noting that it was more intuitive to click an object (as opposed to waiting for the dwell timer). The dwell selection mechanism was reported as slow and at times annoying if the point method did not stay on the virtual object long enough to successfully complete the required 1 second of dwelling time. This was compounded if the virtual element was small and hard to keep the point method engaged in the correct area.

Table 5. Summary of Input Modality Rated as Most and Least Preferred

	Most Preferred	Least Preferred
Controller + Click (*n* = 8)	2	1
Eye gaze + Click (*n* = 4)	1	0
Eye gaze + Dwell (*n* = 5)	0	**5**
Head tracking + Click (*n* = 4)	**4**	0
Head tracking + Dwell (*n* = 4)	1	1
Laser + Click (*n* = 8)	**5**	2
Laser + Dwell (*n* = 6)	1	**4**

Note. The top 2 for each column are identified in bold text.

DISCUSSION

The study reveals that a click selection mechanism is preferred over a dwell selection mechanism. This supports the findings of the previous study (Fussell et al., 2024) as well as the literature (Hou & Chen, 202; Xu et al., 2022) but expands on the use cases by incorporating the eye tracking input modalities. The eye gaze point methods were new modalities within the testbed. The addition came with challenges related to calibration. Many participants struggled to affix the eye tracker to the virtual elements for interaction, while some reported high ease of use. As noted by participant 6448461A, "I preferred the eye tracking [+ click] modality the least because of the accuracy of the eye tracking wasn't always to where I was looking and the calibration wasn't completely accurate." This is reflected in unacceptably high workload scores, especially for the Eye gaze + Dwell modality. For those who struggled, resetting the eye calibration could increase the tracking accuracy. However, this required the researcher to intervene. In a more dynamic training situation, the ability to pause, reset eye calibration, and resume training may not be an option. The difference in sickness scores between the eye gaze modalities also warrants further research to understand if the dwell aspect exacerbates levels of simulator sickness, or if the small group randomly assigned to the modality for this study simply did not encounter symptoms. The complete lack of symptoms for those who used Eye gaze + Click, compared to the high rate of symptoms reported by those who used Eye gaze + Dwell, may indicate that effort required to focus the pupil, wait for the selection cue to appear, and keep the gaze on the item for the set duration (1 sec total) to confirm selection can cause a concerning level of nausea, oculomotor, disorientation, and overall sickness symptoms. Those interested in incorporating an Eye gaze + Dwell input modality into a VE may need to do so for limited interactions.

The higher oculomotor scores for the dwell selection mechanisms also require additional investigation. All scenarios were completed in under 15 minutes. A more complex simulator with a longer training requirement that requires a dwell selection mechanism could lead to exacerbated symptoms, negatively impacting the training outcomes. In training programs that require multiple training sessions or hours-long training, a VE with an input modality that causes or exacerbates symptoms of simulator sickness will lead to ineffective training outcomes and possibly delay the training program as students end training sessions early due to sickness or hesitate to use the VE out of concern.

The participant feedback further confirms the desire for a direct interaction with a virtual environment that is intuitive and not reliant on waiting for a set amount of time to pass. The findings are exemplified by the response of one participant: "I preferred the hand controls with the clicking mainly for the sense that I am clicking the buttons [Controller + Click]. While hovering over the switches made for an interesting experience, I personally prefer to know when I am going to activate a switch. The other two [Eye gaze + Dwell, Laser + Dwell] where [sic] a brand new experience which made the process of the flight sim a little more complex." The ability to reach out to interact with a training environment can promote situated and embodied learning by linking realistic behaviors, interactions, and exploration. This learning can also enhance mental schema that facilitates a transition from the training environment to the real world.

The high sickness and workload scores, coupled with the below average usability score, of the Controller + Click condition may be due to the virtual flight deck panels being raised in the air. It was noted by the researcher and by some of the participants that the panels were higher in the VE than was comfortable to interact with, causing the participants to reach and lean in exaggerated movements as compared to the other input modalities. These scores reflect the importance of VE design and development with ergonomic considerations to ensure a realistic environment that promotes natural movements. Some participants chose to stand to interact with the top panel elements. Extrapolating to a more complex XR simulator, this action is not ideal and could even be restricted by physical elements of the simulator. It may not be enough to rely on starting each user in the same location by relying on the initial set up of the VE when the program is enabled. Rather, the ability to reset the VE to the user's location, or to move elements within the VE to the user (as in, move the virtual desk in the testbed up or down instead of having a fixed location based on room set up) must be considered to ensure optimal interaction and ensure correct, natural behaviors. High sickness and workload scores in a VE—or any simulator—can lead to a negative learning experience, impacting the learner's ability to take in new information, develop mental models, and fully engage in the experience (DeVeaux & Bailenson, 2022; Makransky et al., 2021; Sweller, 2011). This could result in learning undesired behaviors to compensate for the high workload or feelings related to sickness or a negative transfer of training.

Although statistical analyses were inconclusive, a future round of data collection, to reach the required sample size of 43, and equal distribution among the input modalities is warranted. While there is a clear preference based on open-ended feedback for selection mechanisms (i.e., click and dwell), the E-SUS, SSQ, and NASA TLX scores do not trend in the same manner. Analysis with larger sample sizes may lead to further understanding of the extent to which each input modality impacts usability, sickness, and workload, and how these elements influence preference.

FUTURE RESEARCH

Although data were collected to answer research question 1, it was not analyzed for the present report. A future report will analyzed measures of efficiency and effectiveness, logged within the program based on participant activity. Efficiency will be measured as the time to complete an operation and task, the time between operations and tasks, and the time to complete a scenario. Effectiveness will be measured by error rate and number of attempts to accomplish an operation.

In the near-term, the VE will receive an update to adjust the height of the virtual elements to the user and refine other issues that were revealed. Development will continue to enhance eye tracking, including dead zones around virtual components requiring a threshold to trigger movement, the ability to lock onto an object, and movement smoothing. Further data collection is planned to increase the sample size of the seven input modalities to 43 and conduct statistical analyses. The logging data will also be analyzed to compare the efficiency and effectiveness of the input modalities.

The VE will also be expanded to enhance the complexity of the testbed by setting it within a virtual flight deck as opposed to an office space. Future development includes the design and testing of an MR modality as well as a control (i.e., interaction with a physical flight deck) to investigate the interaction of levels of immersion with usability, sickness, workload, and preference. A future study will measure "immersion" as a binary condition with full immersion (VR) or structured interaction with real-world and virtual elements (MR). Expanding the testbed to include additional virtual elements with realistic procedures will allow for more generalization of the findings beyond a simple, procedural trainer while also exploring new ways to incorporate different input modalities into a VE.

ACKNOWLEDGEMENTS

The authors wish to thank 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. Thanks are also given to Kent State University College of Aeronautics and Engineering for partnering in data collection.

CLEARANCE

The views, opinions, and/or findings contained in this journal article are those of the authors and should not be interpreted as representing the official views or policies, either expressed or implied, of the Air Force Research Laboratory, Department of the Air Force, or the Department of Defense. Publication is cleared, AFRL-2025-2945.

REFERENCES

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies, 4*(3), 114-123. https://uxpajournal.org/determining-what-individual-sus-scores-mean-adding-an-adjective-rating-scale/

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, *189*(194), 4-7.

DeVeaux, C., & Bailenson, J. (2022). Learning about VR in VR. *XRDS: Crossroads, The ACM Magazine for Students, 29*(1), 14–19. https://doi.org/10.1145/3558189

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

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., Oliver, Q., Frost, T., & Kwasa, B. (2024). Development and testing of extended reality input modalities for a virtual learning/training task. *Proceedings of the 2024 Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).* Orlando, FL.

Harper, S. B., & Dorton, S. L. (2021, September). A pilot study on extending the SUS survey: Early results. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 65, No. 1, pp. 447-451). Sage CA: Los Angeles, CA: SAGE Publications.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139-183). North-Holland.

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

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, *3*(3), 203-220.

Kennedy, R. S., Drexler, J. M., Compton, D. E., Stanney, K. M., Lanham, D. S., & Harm, D. L. (2003). Configural scoring of simulator sickness, cybersickness and space adaptation syndrome: Similarities and differences. In *Virtual and Adaptive Environments* (pp. 247-278). CRC Press.

Makransky, G., Andreasen, N. K., Baceviciute, S., & Mayer, R. E. (2021). Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. *Journal of Educational Psychology, 113*(4), 719–735. https://doi.org/10.1037/edu0000473

Rebensky, S., Stalker, W., Diemunsch, J., Turk, S., Oliver, Q., Chase, M., Crawford, E., & Rodabaugh, T. (2023). *Current evolution of VR input modalities*. *Proceedings of IT2EC 2023*. Rotterdam, the Netherlands.

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/

Sauro, J. (2018, September 19). *5 ways to interpret a SUS score*. Measuring U. https://measuringu.com/interpret-sus-score/

Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation* (Vol. 55, pp. 37-76). Academic Press.

Urban, D. & Pritchard, R. (2024, December). Beyond illusions: Navigating VR fidelity in undergraduate pilot training - A 3-year data analysis. *Proceedings of the 2024 Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).* Orlando, FL, US.

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