Virtual Reality (2026) 30:8
https://doi.org/10.1007/s10055-025-01263-5

SI.: PROTOTYPING AND DEVELOPING REAL-WORLD APPLICATIONS FOR EXTENDED REALITY

Benefits of virtual reality and 3D visualizations for remote satellite supervisory control

Savannah Buchner[1] · Jennifer Wood[1] · Gabriella Schauss[1] · Blake Bateman[1] · Hanspeter Schaub[1] · Allison Hayman[1]

Received: 31 January 2025 / Accepted: 30 October 2025

Abstract

Remote supervision is becoming an increasingly complex and important paradigm for satellite operations, especially as on-orbit servicing, maintenance, and assembly are becoming feasible. The limitations of current 2D dimensional displays, such as difficulties interpreting data, may present challenges for these operations. However, the use of virtual reality (VR) shows promise for direct control paradigms and has been positioned to provide benefits for other remote supervisory paradigms. This paper investigates the effects of immersive VR displays and 3D visualization for a display for a simulated satellite rendezvous mission. Three different displays are compared through human subject testing (n = 45) on performance, situation awareness (SA), workload, usability, and subjective utility. These displays include an immersive VR display, a computer display with 3D visualizations, and a traditional computer display with no 3D visualizations. The results show that VR and 3D visualizations improved performance on hard trials (p = 0.008), and had higher perceived utility (p = 0.006). No differences were found in participant's SA, workload, and perceived usability. This research indicates that for remote satellite operations, it may be beneficial to include 3D visualizations, but that the additional inclusion of VR-based immersion may not significantly improve operations.

Keywords Virtual reality · Situation awareness · Human factors · Supervisory control

1 Introduction

Remote supervision, in which a human operator monitors and sends intermittent commands to a system, can be a challenging yet important control modality for operation and exploration tasks, like spaceflight, transportation, or manufacturing. The majority of an operator's time is spent monitoring, which is critical for understanding system states, anticipating potential issues, or quickly detecting failures (Sheridan 2012). Additionally, operators may need to send intermittent commands to avoid potential issues before they occur. Many future supervisory operations will likely be remote, where operators and the systems they work with are separated spatially and potentially temporally (i.e., due to time delay of sending information), like in a spacecraft mission control center. Remote operations can be challenging due to the lack of environment context and the decoupling of processing abilities from the physical environment (Chen et al. 2007). This lack of context can create additional challenges for displaying information, including compromised situation awareness (SA), and reduce the mission completeness (Chen et al. 2007; Endsley 2000). SA is defined as having 3 levels where level 1 is "the perception of critical elements in the environment", level 2 is "the comprehension of their meaning" and level 3 is "the projection of their status into the future" (Endsley 1988).

Satellite operations is a supervisory operational domain which is often understudied. Beyond the challenges present in supervisory control, future proposed mission objectives may present additional difficulties. For example, on-orbit inspection and servicing tasks are becoming increasingly desirable, but involve multiple satellites in close proximity. The relative trajectories of the satellites must be quickly understood by operators in order to avoid collisions, but are governed by flight dynamics that can be complex and

---

✉ Savannah Buchner
savannah.buchner@colorado.edu

[1] Ann and H.J. Smead Department of Aerospace Engineering Sciences, University of Colorado Boulder, Boulder, CO 80303, USA

non-intuitive. Additionally, the close proximity between satellites during these tasks results in less time to make decisions and send commands. Critical events, such as for collision avoidance maneuvers, may necessitate actions to be made in seconds (Sellmaier and Frei 2022). The complexity of the system increases with uncertainties in the states of the satellites and can be especially significant when repairing a dead or injured satellite.

Current satellite operations take place in a mission control room. While different organizations have different protocols and displays, the majority of displays consist of densely packed telemetry data, 2D graphical representations of values over time, or an in-plane representation of orbital motion (Sim et al. 2008). Mentally processing 3D data that is represented in 2D can increase the operator's workload (Bualat et al. 2013; Dan and Reiner 2017; Woods 2011). Workload can be influenced by the task and environment, as well as the operators' skills, behavior, or perception, and is defined to be the cost incurred by an operator to achieve a particular performance level (Hart and Staveland 1988). Additionally, workload can be related to the amount of data needed to be processed (Warm et al. 2008), and having to quickly comprehend and react to the data can further increase workload. To improve these displays for next-generation supervisory control rooms, it has been suggested to develop systems to manage the large amounts of data and provide only needed or requested information, which would improve SA and reduce workload (Sim et al. 2008).

Virtual Reality (VR) has been proposed to help improve control rooms and operations in a variety of supervisory control situations like spaceflight (Sittner et al. 2023; Gad et al. 2023), maritime (Lager and Topp 2019; Tsigkounis et al. 2021) and air traffic control (Gorbunov and Nechaev 2022; Cordeil et al. 2016). For satellite operations in particular, VR has been suggested as a way to improve environmental context and allow operators to better understand 3D orbits. VR increases telepresence, which is the feeling of being in an environment other than where one is physically (Sheridan 1992), and can provide environmental context. This is due to the immersive nature of VR, where immersion is defined as "the extent to which the computer displays are capable of delivering an inclusive, extensive, surrounding and vivid illusion of reality to the senses of a human participant" (Slater et al. 1996).

While supervisory control is relatively understudied for VR operations, there has been a lot of research into monitoring and manual control operations with VR. For spaceflight monitoring, previous work has compared VR, 3D screen visualizations, and a traditional 2D display. This found that VR and 3D visualizations improved level 2 and 3 SA over a traditional display, but that there were no differences in workload or usability (Buchner et al. 2025). However, this task was only monitoring and participants had no control authority. Similar display types studied for a maritime monitoring task found that VR and 3D screen visualizations improved SA and collision detection (Lager and Topp 2019). In a self-driving car monitoring study it was found that VR decreased usability, and increased workload and simulator sickness (Kalamkar et al. 2023). The authors of that study acknowledge this was likely due to a combination of the hardware and not designing or optimizing the display for VR. Additionally, an air traffic control application study found VR reduced the number of errors made in identifying dangerous situations compared to traditional computer displays (Gorbunov and Nechaev 2022). As supervisory control requires more than just monitoring it is unclear how these results translate.

Beyond the previous monitoring studies, VR has been heavily studied for manual control tasks. During these VR has been found to improve depth perception and collision avoidance, lead to faster task completion, increase the sense of presence, improve usability, and reduce the perceived effort compared to 2D displays (Chen et al. 2007; Tittle et al. 2002; Elor et al. 2021; Whitney et al. 2020a). There have been mixed results on whether VR can improve SA or reduce workload (Elor et al. 2021; Whitney et al. 2020b, a; Hosseini and Lienkamp 2016; Read and Saleem 2017). Manual control tasks require different cognitive demands and may benefit from different aspects of VR than supervisory control tasks. As such, we cannot assume that these benefits of VR will translate to supervisory control tasks.

While VR has potential benefits for supervisory control tasks and satellite operations, it also has some limitations that may lead to challenges for use as an operational display. Some of these limitations are due to current hardware and may resolve themselves as technology advances. Current VR technology lacks the resolution to display large blocks of text, like those from traditional monitoring displays, and instead has to rely more on visual representation of the data (Van de Merwe et al. 2019). Additionally, operator buy-in and susceptibility to cybersickness can be influenced by display design choices. These include aspects such as viewpoint selection, field of view, and amount of control over the environment (van Emmerik et al. 2011; Davis et al. 2014). This can additionally be influenced by properties inherent to the headset, such as resolution or weight. If an operator finds it uncomfortable or hard to use, it will limit its effectiveness as a display.

The objective of this research is to compare displays with different degrees of immersion or 3D visualization on a satellite supervision task where subjects will have some, but limited, control authority over their satellites. This includes an immersive VR display, a 3D visualization computer-based display, and a baseline display with 2D graphical representation. These will be compared on measures of performance, SA, workload, usability, and subjective utility. We hypothesize that 3D visualizations and VR will lead to higher performance, SA, usability, and subjective utility, over the baseline display. We further hypothesize that VR will lead to an increase in these measures over the 3D visualization display. Finally, we hypothesize there will be no differences in workload among the displays.

2 Methods

In this research, three displays are designed to simulate a remote supervision of satellite operations. The simulated scenario is a spacecraft inspection task, in which an operator assists a servicer satellite to perform corrective burns and inspect a client satellite. These displays are compared through a human subject evaluation.

2.1 Scenario

Participants were tasked with monitoring and supervising the proximity portion of a satellite rendezvous mission. The underlying trajectories used for the simulation were developed using Basilisk,[1] a high fidelity, flight-proven, physics-based satellite simulation tool (Kenneally et al. 2020). To enable evaluation in a laboratory environment, the simulation is sped up by a factor of 15. The scenario (Fig. 1) consists of two satellites in orbit around Earth: a non-operational, tumbling, client satellite and an active servicer satellite, supervised by a remote operator, sent to inspect the client satellite. The goal of the operator is to successfully service the client or abort, if necessary, to avoid a collision. The client satellite has no communications, fuel, or battery, thus there is uncertainty in its location and attitude. In addition, there is uncertainty in the velocity change imparted by the thruster burn. These uncertainties are combined and visualized as a spheroidal keep out zone, which the servicer should avoid entering or there may be a collision. The keep out zone grows and shrinks based on environmental factors and operator actions. A proximity sensor is simulated to have better performance when the satellites are in the sunlight, causing the keep out zone to shrink; when they are in the shadow it grows due to worse sensor performance.

The servicer satellite begins on a drift orbit passing by the client satellite. During the simulation, the servicer fires its thrusters to change its relative orbit to circle the client satellite. The operator can select from 1 of 3 burn locations, each separated by 15 min of flight time. The ideal burn location selection may be influenced by the lighting conditions and battery levels. Different burn locations may result in successfully servicing the client satellite or being required to abort. The operator can also turn on a light, which mimics using a better sensor, to reduce the uncerftainty; however, this also drains the battery at a faster than nominal rate. The battery nominally decreases slowly in the darkness and increases in the sunlight. If the servicer enters the keep out zone, this would count as a collision. If a collision is unavoidable, the operator can choose to abort the mission provided they meet fuel, battery, and time criteria. Using the onboard light too much may result in an abort being impossible due to a low battery.

Different scenarios are created by manipulating the initial orbit, date, lighting conditions, keep out zone size, fuel, and battery levels. This influences the outcomes of which burn location is ideal, how long the light can be turned on for before the battery is depleted, and if the operator is forced to abort. Each trial lasted up to 7 min. Aborting or colliding would cause the trial to end earlier. Eight experimental scenarios and two familiarization trials were developed. The experiential scenarios were of varying difficulty levels, grouped into Easy, Medium, and Hard. The difficulty was based on the number of specific decisions the participant had to make correctly for a successful trial. For easy trials, multiple burn locations and a wide range of light use durations and times would lead to success. In Medium trials, one or two burn locations would lead to success, but the light must be used in a more restricted manner (i.e., a smaller

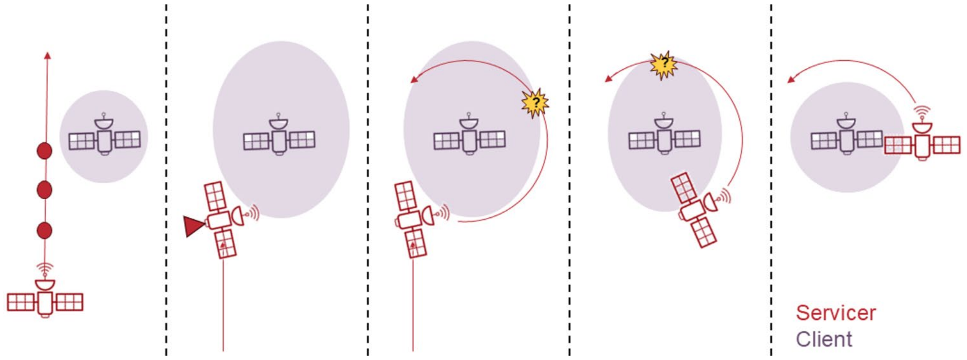


Fig. 1 A 2D diagram of the scenario. From left to right: the servicer satellite on the drift orbit with 3 potential burn locations; the servicer making a burn, increasing uncertainty; the propagation of this orbit leads to a potential collision; the keep out zone reducing with time and light; there is no collision risk

---

[1] https://avslab.github.io/basilisk.

range of use times or durations is required). Finally, on Hard trials only a specific burn location and an even more constrained use of light is required for success. Hard trials were designed such that the most likely outcome was an abort.

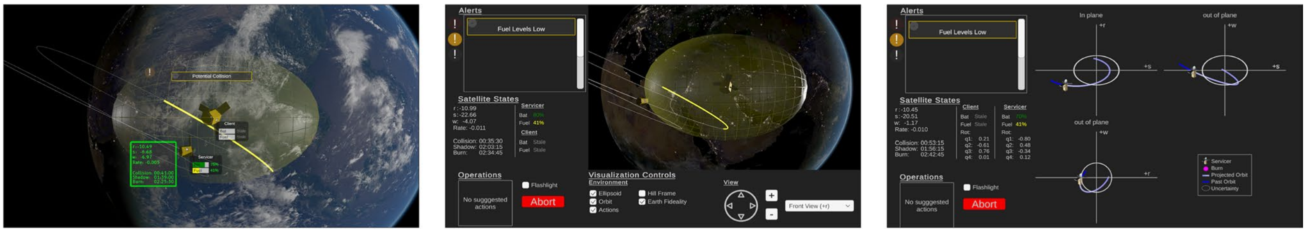
2.2 Display designs

Three displays were designed for this experiment: a VR display, a 3D screen visualization (Scr. Viz.), and a two-dimensional Baseline display as seen in Fig. 2. These displays were based on those developed in Buchner et al. (2025) but with modifications to allow user input and selection. For this experiment, the VR display was controlled using a Meta Quest 2 headset and controllers, while the other two displays were designed for a 2D monitor and mouse. The displays were built using Unity 2022.3.21f. All three displays present the same information to the operator, but that information is conveyed with different degrees of immersion and 3D visualizations.

The VR display is seen in Fig. 2a. The underlying visualizations, including satellite models, relative orbital motion, accurate Earth models, and appropriate sun-based lighting were based on the Vizard spacecraft simulation visualization software application (Wood et al. 2018). Overlaid were visualizations of the keep out zone, centered around the client satellite, and colored based on collision risk. Red represented a warning, with less than 15 min to a potential collision, yellow represented a potential collision at any point within the next two relative orbits, gray represented no collision risk, and blue indicated the keep out zone was at a minimum size. The portion of the orbit line where it entered the keep out zone was changed to the corresponding condition color. Finally, a representative light was visible when the participant turned the light on.

Beyond visualizations, critical information and user input options were displayed in a text-based form. A heads-up display (HUD) displayed the range, rate, and time to burn, collision, and lighting changes. This was always located in the same peripheral location of the headset allowing participants to access information through eye movements, minimizing information access effort (Wickens 1993) and minimizing blockage of the visualization. Its location and text size were based on VR recommendations (Yao et al. 2014; Mckenzie and Glazier 2017) and adjusted based on feedback during developmental human factors testing. Satellite states, such as fuel and battery, were presented both as text and a gauge and moved with their associated satellite. This was done to minimize the text in the HUD and help preserve the immersion (Iacovides et al. 2015; Rosyid et al. 2021; Marre et al. 2021; Caroux and Isbister 2016). Cautions and warnings were displayed at the top of the screen and designed to be easily noticed by an operator. These were triggered automatically for certain events, including low battery, fuel, and potential for a collision. These alerts could be minimized to remove them from the primary field of view or maximized to review them again (O'Hara and Fleger 2020) , (2004). Finally, the user input was provided through a panel with various options presented above the HUD when available. User input panels disappeared after an option was selected or the time window for user input to the panel closed.

The participants could control their visualization and operations through a radial menu system and their hand controller as seen in Fig. 3. This allowed them to turn on and off visualization or text components. Additionally, they could change their viewpoint or perform operations like view alerts, turn on the light, or abort. The right controller was used to control the menu and user inputs. Beyond the menu, participants could interact through their hand controllers. The left-hand controller controls navigation, such as panning, zooming, and teleportation. This allows users to explore the satellite operational change and easily change their viewpoint. In conjunction with this, participants could navigate the scene through body movements. Participants had access to a 10 by 10 foot area of floor to walk around; the tracking dynamics were set such that the participant could cover the distance to the satellite while walking around it in this area. Beyond this, head or body rotation allowed the user to change their perspective quickly. Additionally, preset buttons allowed easy access to turning on/off the light or to abort. An abort required confirmation to ensure it was not selected in error. The display and controls underwent human factors testing prior to the experiment to ensure the text was readable and the controllers were acceptable.



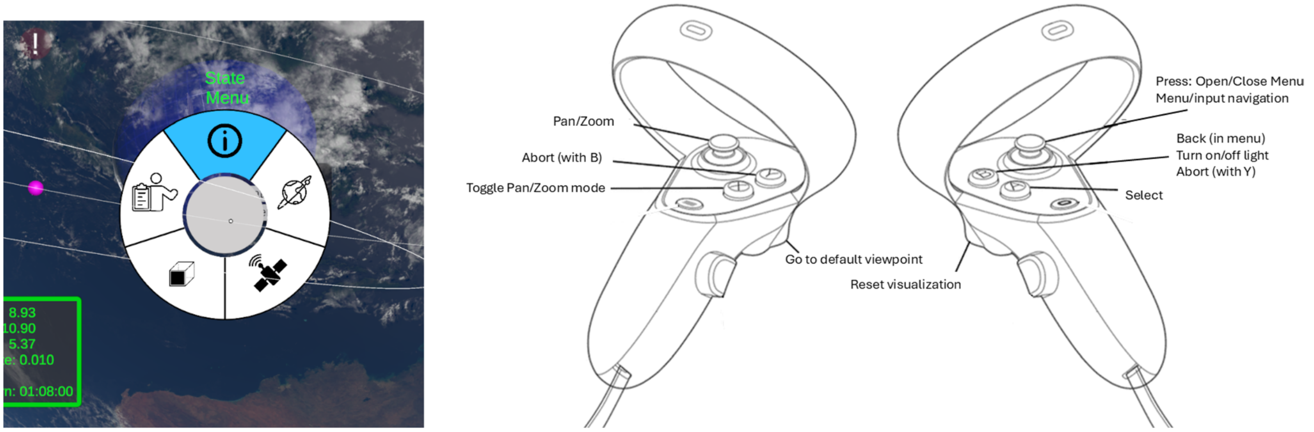
r: -10.49
s: -6.68
w: -6.97
Rate: -0.005
Collision: 00:41:00
Shadow: 01:39:00
Burn: 02:25:30

r: -10.99
s: -22.66
w: -4.07
Rate: -0.011
Collision: 00:35:30
Shadow: 02:03:15
Burn: 02:34:45
Servicer Bat 80% Fuel 41%
Client Bat Stale Fuel Stale
☑ Ellipsoid
☑ Orbit
☑ Actions
☐ Hill Frame
☑ Earth Fidelity

r: -10.45
s: -20.51
w: -1.17
Rate: -0.010
Collision: 00:53:15
Shadow: 01:56:15
Burn: 02:42:45
Client Bat Stale Fuel Stale Rot: q1: 0.21 q2: -0.61 q3: 0.76 q4: 0.01
Servicer Bat 70% Fuel 41% Rot: q1: -0.80 q2: 0.48 q3: -0.34 q4: 0.12

(a) (b) (c)

Fig. 2 The three different display designs. **a** VR display **b** Screen Visualization display **c** Baseline display



(a) (b)

Fig. 3 User navigation and controls in VR **a** The radial menu; the example image shows the top level menu in a zoomed in view of the VR display, **b** The Oculus hand controllers mapping

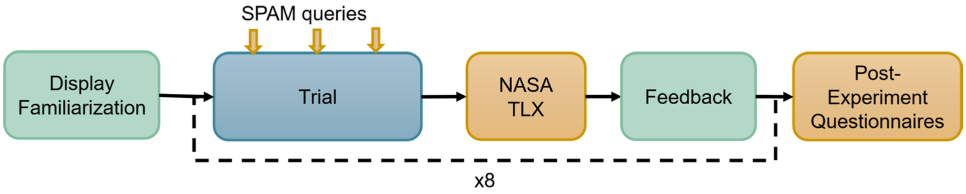


Fig. 4 Experimental design flowchart. The orange boxes indicate data collection through surveys or queries

The Scr. Viz. display maintains the 3D visualization described for the VR display but does not include the immersive environment that VR does. This can be seen in Fig. 2b. The operator can still interact with the system by panning, zooming, and customizing their visualization but this time on a 2D computer screen. All text, including the HUD, satellite states, alerts, and user input are displayed on the screen outside the visualization, allowing for a consistent scan pattern.

Finally, the Baseline display was designed to be consistent with the text-heavy and graphical displays currently used for most traditional satellite operations. It has only 2D visualizations where the 3D visualization is replaced by graphical telemetry. The range is displayed as plots of the different orbital planes and past and future motion. The same text-based interface in the Scr. Viz. display also surrounded the telemetry plots. For both the Scr. Viz. and the Baseline displays, navigation, and interactions were done with a mouse.

2.3 Experimental design

The study was approved by the University of Colorado Institutional Review Board (#24-0250). Informed consent was obtained from all participants. 45 participants from around the University of Colorado Boulder campus were enrolled and completed data collection (18 female, 27 male; ages 18-38, median age 23 years; 25 with prior orbital experience; 15 with prior satellite operational experience; 21 graduate students, 17 undergraduate students, 5 with complete undergraduate degrees). The participants have similar background and education as the type of people that operate satellites. 47 participants were enrolled, but 2 voluntarily did not complete the study and thus were removed. Participants were aware of the high-level project goals from the informed consent but naive to the exact manipulations or alternative display designs. Participants were screened to ensure their vision was correctable to 20/20, they were not colorblind, and they scored less than 90% on the Motion Sickness Susceptibility Questionnaire (Golding 1998). This was used to identify individuals who may be highly susceptible to simulator sickness before data collection. Additionally, participants in VR were monitored for cybersickness symptoms, particularly nausea, throughout the experiment; there were no reported cases of nausea in VR.

Three different displays were compared for satellite operations: the VR, Scr. Viz. , and a baseline display as described above. Participants were randomly assigned to one of the three displays for the task and followed the procedures as described below and in Fig. 4.

During their visit, participants were first familiarized with the task and their assigned display. This was done through a PowerPoint which reviewed the task's motivations and goals, as well as how to use their display. They then did three familiarization trials. The first trial was primarily an opportunity to get used to the controls and display—there was no collision risk present. The next two familiarization trials mimicked a real trial and had identical initial states to give participants an opportunity to execute different actions so they could understand the impact on the mission outcome. This was done to ensure that participants understood how to use the displays and perform the task before the real trials began and to eliminate issues of performance due to lack of familiarity with the display. During both the PowerPoint training and training trials subjects were quizzed to ensure they understood the necessary information.

Participants then completed eight trials. Every participant completed the same eight trials, but the scenarios were presented in a randomized order. During each trial, SA, workload, and performance were assessed. To assess the three levels of SA, the Situation Presence Assessment Method (SPAM) was used (Durso et al. 1998). This is a real-time SA assessment method that has been used in other operational-based experiments, such as air traffic control rooms and maritime/submarine control rooms (Durso et al. 1999; Loft et al. 2013, 2015; Cunningham et al. 2015; Mirchi et al. 2015; Fujino et al. 2020). It is designed to mimic a control room setting where operators may be asked questions but still have access to their display to respond. Up to three points throughout the trial, an auditory tone was played. Participants were instructed to provide verbal confirmation when they had the ability to answer the questions. At this point, or after 20 s had elapsed, an experimenter asked the participant three questions, one per SA level. Questions were generated through a process similar to a goal directed task analysis (Endsley 2000). Example questions include: “Is the servicer satellite currently in the sun? (level 1)”, “Is the portion of the orbit line in the keep out zone decreasing? ” (level 2), or “If no new action is taken, will the keep out zone be shrinking in 30 min?” (level 3). The full list of questions is in the supplementary materials. Participants were instructed to respond as quickly and as accurately as possible and were allowed to use the display to answer. If a trial ended early due to an abort or collision, fewer questions may have been asked.

After each trial, participants assessed their workload through the NASA Task Load Index (TLX) (Hart and Staveland 1988). To do so they rated six dimensions of workload on a 21-point scale. This includes mental, physical, temporal, performance, effort, and frustration. At the end of all eight trials, participants then completed comparisons between the different components, allowing for a weighted workload score to be calculated. After completing the TLX survey for a trial, participants were given feedback on their performance. This included information about their burn performance (did they make a good burn selection that could lead to success), end state performance (considering success, abort, or collision), and combined total performance. Each of these was on a scale of 'poor’, 'fair’, 'good’, and 'excellent’. This performance also corresponded to a monetary bonus participants could earn between $\$-1.00$ and \$1.00 per trial.

Table 1 Summary of dependent variables and measurement techniques

Dependent variable	Measurement technique
Performance	Ordinal outcome score per trial, separated by trial difficulty
SA	Average of SPAM queries answered correctly on a per level basis
Workload	Weighted NASA TLX per trial
Usability	SUS
Subjective utility	Custom questionnaire

Finally, after all trials were complete, participants completed the System Usability Scale (SUS) (Brooke 1996), which is a 10-question survey in which participants respond on a 5-point scale. Their answers to each question are combined to give a resulting score from 0 to 100. In addition, participants answered a custom subject utility survey which had Likert-style questions relating to their perceived understanding of the events, uncertainties, collision likelihood, orbital motion, and operational decision, as well as free-response questions about their experience. The full survey is provided in supplementary materials. They also completed a demographics survey including information about sex, orbital mechanics familiarity, operational familiarity, and VR familiarity. Their familiarity was coded as a binary little to none, or moderate to high. Finally, they completed the balloon risk analog task (BART) (Lejuez et al. 2002; Basner et al. 2015) to assess their risk-taking behavior.

2.4 Statistical analysis

Analysis was performed to compare the three displays on performance, SA, workload, usability, and subjective utility, summarized in Table 1. The study collected 360 trials over 45 participants. One score was collected for each subject for SA level, usability, and utility, for a total of 45 data points per assessment type. For workload and performance, trials were kept separate and so the 360 data points were used. A criterion of $\alpha = 0.05$ was used for significance for all tests, and all assumptions for each statistical test were met.

Using the results of a previous satellite monitoring study (Buchner et al. 2025), sample size calculations done with G*Power indicated that 15 participants per group would be sufficient at a 80.1% power to detect differences using an one-way ANOVA with $\alpha = 0.05$ for all hypothesized differences.

Prior to statistical analysis, the SA, workload, and performance data were inspected for a confounding factor of trial order to capture undesirable learning effects. No effect of trial order was identified for any analysis. However, for workload and performance, there was a dependence on the specific scenario (i.e., regardless of the order presented, some trials had consistently different workload and performance than others). Thus, for these statistical tests, these factors were included in the model. In addition, all data was evaluated for confounds based on participant background including: sex, orbital mechanical familiarity, satellite operational experience, VR familiarity, and BART score. The participant background metrics that were evaluated were selected as they may influence how participants respond (i.e., being more familiar with orbital mechanics may result in better SA performance, an increase in risk tolerance may lead to different performance, or sex may lead to differences in how displays were used). These were included in the models as appropriate after visual inspection of the data.

Performance is on a 12-point ordinal scale, and as such a cumulative linked mixed model approach was taken (using clmm in R (Christense 2023)). The performance score is a ranked score based on the participant's decisions on the burn location and end state. The end state refers to the final state of the servicer (collided, aborted, or serviced). Granularity is added to the score by evaluating the participant's abort attempts (was the abort necessary or unnecessary at that time), use of the light, and end battery level. These different actions and metrics provide indirect information about the participants' understanding of the task, awareness of the environment, and optimization of the outcome. The full metric calculation is described in the supplementary material. The training display and its interaction with trial difficulty were included as a fixed effect, and the number of balloons collected in the BART task (Reed et al. 2012), which has been shown to be correlated with risk, was also included as a random effect. We assessed this model, but there was a significant interaction, indicating that display modality may not provide differences across difficulty levels. Thus, we separated the data by scenario difficulty level and analyzed them separately, keeping the same fixed and random effects. A type III ANOVA was used to compare performance across conditions (anova.clmm in the RVAideMemoire package in R (Maxime HERVE 2011)). Post-hoc tests were done between all pairwise comparisons using estimated marginal means (Lenth et al. 2023) with a Tukey-*p*-value correction.

As another measure of performance, the overall outcomes were compared across training conditions, including the number of trials aborted, the number of trials resulting in a collision, and the number of successful trials. This was done using a Kruskal-Wallis one-way analysis of variance.

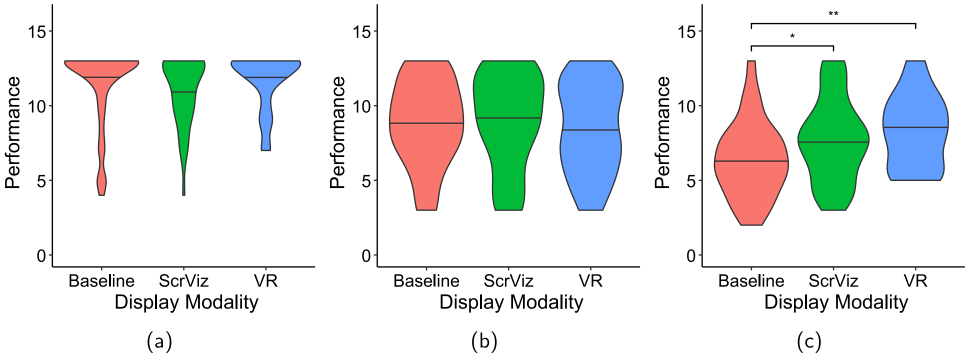
Each SA level was analyzed independently using the percent of queries answered correctly across all trials. This resulted in one measure per participant per SA level. This transformation into a percentage was made because the participants answered different numbers of questions due to their individual performance (i.e., aborting a trial may result in less SPAM queries answered). For each SA level, a linear mixed effects model was used to compare the effect of display modality on SA. The display modality was considered as a fixed effect. The random effects were dependent on SA level, with level 1 SA including orbital mechanics familiarity and sex, while level 2 and 3 SA included both operational and orbital mechanics familiarity. The model was fit using the lme4 package in R via penalized maximum likelihood estimation (Bates et al. 2015). The significance of display modality was assessed using an F test with a type III ANOVA with a Satterthwaite approximation for degrees of freedom and was implemented using lmerTest package in R (Kuznetsova et al. 2017). Any necessary post-hoc tests were done between all pairwise comparisons using estimated marginal means (emmeans package in R (Lenth et al. 2023)) with a Tukey p-value correction and Kenward-Roger degrees of freedom correction. Effect sizes were calculated using the effectsize package in R (Ben-Shachar et al. 2020)

For workload, the unique weighted TLX score for each trial was analyzed, resulting in 360 total data points. As with SA, a linear mixed effects model was fit using the training display condition as the fixed effect. The participant and scenario were included as random effects to account for the repeated measures and differences across each scenario. Likewise, for usability a linear mixed effects model was fit with the training display as a fixed effect and sex as a random effect. In both cases the same pipeline as SA was followed.

The subjective utility questions were each on a 5-point Likert scale. Thus, ANOVAs could not be used to compare ratings across conditions, and instead, each question was analyzed using a Kruskal-Wallis one-way analysis of variance. Post-hoc tests were done using Dunn's test with a Holm correction (Alexis 2014).

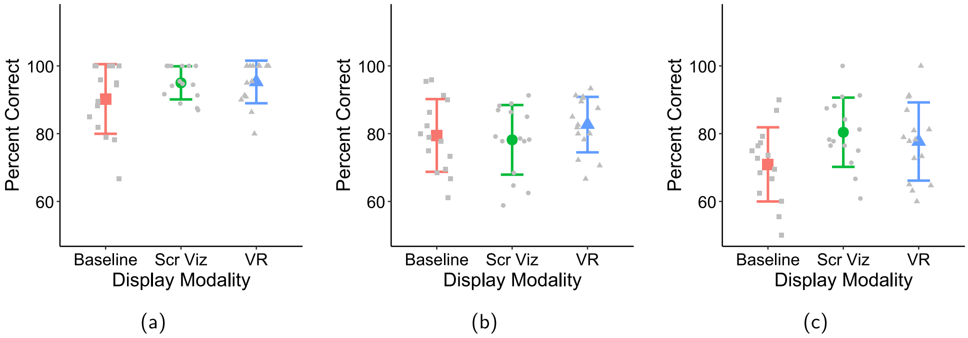
3 Results

In the comparison of participant's performance, significant differences were found among the Hard scenarios ($\chi^2(2) = 9.61, p = 0.008$), but no differences among the Easy ($\chi^2(2) = 3.53, p = 0.17$) or Medium ($\chi^2(2) = 0.70, p = 0.70$) difficulty scenarios. Note that for both the Easy and Medium conditions, subject performance was frequently at the maximum of the scale. For the Hard scenarios, differences were seen between Baseline and VR ($z = -3.06, p = 0.006$) and Baseline and Scr. Viz. ($z = -2.36$, $p = 0.047$). For both cases, participants in the Baseline condition performed worse. No differences were found between VR and Scr. Viz. ($z = -0.74$, $p = 0.74$). This can be seen in Fig. 5. No differences



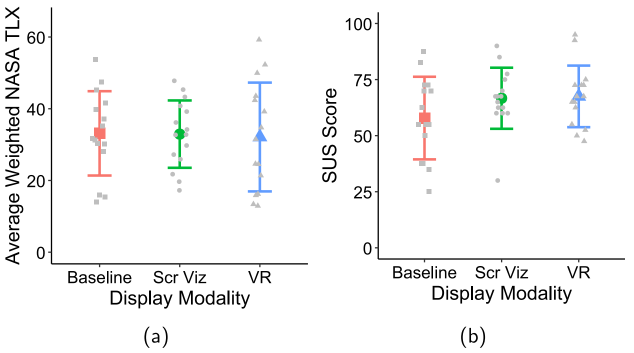
(a) (b) (c)

Fig. 5 Participant's performance on **a** Easy, **b** Medium, **c** Hard trials. The violin plot is overlayed with the median score, and significance is noted between the conditions



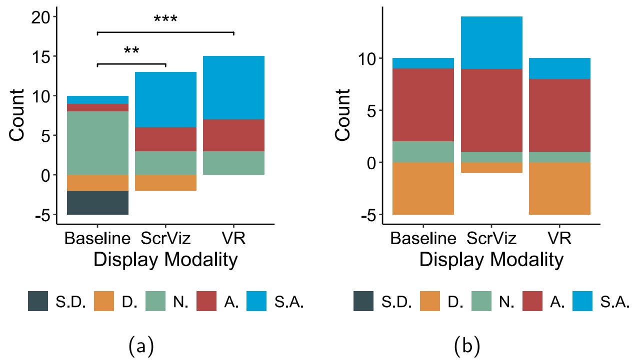
(a) (b) (c)

Fig. 6 The **a** level 1, **b** level 2, and **c** level 3 SA results. All figures show the subject averaged percent of SPAM queries of that level answered correctly. The data mean and standard deviation error bars and significance is noted between the conditions. Note that the Y axis ranges from 50 to 100%



(a) (b)

Fig. 7 The **a** workload, and **b** usability results. Workload shows the subject-average weighted TLX score, and usability shows the System Usability Scale score. The data mean and standard deviation error bars and significance is noted between the conditions



(a) (b)

Fig. 8 The subjective utility questions, showing the results of **a** "I found this system enabled me to understand the relative orbital motion of the satellites" and **b** "I found this system allowed me to make appropriate operation decisions" (S.D. = Strongly Disagree, D. = Disagree, N. = Neither Agree nor Disagree, A. = Agree, S.A. = Strongly Agree)

were found in the number of scenarios aborted (H(2) = 0.39, $p = 0.82$, $\eta^2 = -0.038$), the number of scenarios with collisions (H(2) = 4.20, $p = 0.13$, $\eta^2 = 0.05$), or the number of successful scenarios (H(2) = 0.2, $p = 0.87$, $\eta^2 = -0.04$).

No significant differences in SA were found among the displays in any of the levels, as seen in Fig. 6. The ANOVA comparison between the linear mixed effects models found that all were trending towards, but did not reach, significant differences in the percent of questions answered correctly with Level 1 (F(2,23.11) = 2.3, $p = 0.12$, $\eta^2 = 0.17$), Level 2 (F(2,41.83) = 2.03, $p = 0.14$, $\eta^2 = 0.09$), and Level 3 (F(2,41.57) = 2.42, $p = 0.10$, $\eta^2 = 0.13$).

For workload, no significant differences were found in the ANOVA (F(2,41.7) = .03, $p = 0.97$, $\eta^2 = 0.001$) as seen in Fig. 7a. Additionally, no significant differences were found in usability's ANOVA (F(2,42) = 1.82, $p = .17$, $\eta^2 = 0.08$), as in Fig. 7b. However, the coefficient for the VR term in the linear model approached significance ($p = 0.09$), which indicates that there are trending differences between the VR and Baseline usability score.

Subjects perceived differences in utility between the display conditions in two aspects, as seen in Fig. 8 . Differences were reported in the participant's perceived ability to understand orbital motion (H(2) = 14.6, $p = 0.006$, $\eta^2 = 0.30$) and ability to make operational decisions (H(2) = 6.11, $p = 0.047$, $\eta^2 = 0.10$). Post-hoc comparisons for orbital motion using the Holm correction found differences in the

Baseline and Scr. Viz. ($z = -2.86$, $p = 0.008$) and Baseline and VR ($z = -3.61$, $p < 0.005$) with the Baseline display rated significantly worse in both cases. No differences were found between Scr. Viz. and VR ($z = -0.75$, $p = 0.45$). Additionally, no significance post-hoc comparisons were found for the ability to make appropriate operational decisions. However, the Baseline to Scr. Viz. comparison was trending significant ($z = -2.31$, $p = 0.061$). No other differences in subjectively reported utility were found.

Participants' subjective written comments regarding aspects relevant to the display and/or interface were analyzed. The VR display was found to be polarizing. Some participants enjoyed it *"It is rather intuitive and easy to understand", "It is a good user interface"*, while others found it hard to use *"I felt like there was no benefit to being in VR and it just makes it clumsier and harder to use the menu"*. No participants left insight about the Scr. Viz. display itself. For the Baseline display comments were negative and uniformly indicated a dislike of the 2D nature *"Would be much easier if we could see the orbit paths in 3D vs on a 2D plot", "I had to visualize the 3D space ... which made the task more demanding"*. Similar comments were echoed among the other Baseline participants.

4 Discussion

This study compared displays with different degrees of 3D visualizations and immersion in a remote supervision task and is one of the first studies to investigate VR for remote supervision operations. This research helps fill the gap in understanding of how VR may be used in an operational environment that is not full manual control or passive monitoring.

We found that performance was improved by the use of 3D visualizations, both on the screen or in VR, but only in the trials of the hardest difficulty. This result highlights a benefit of including 3D visualizations, either from VR or on a computer screen. This scenario is already a simplified spaceflight operation for the purposes of research. Thus, these promising results may indicate that as real-world scenarios are implemented, the potential benefit of the VR display could be further enhanced. As satellite operations become more complex and challenging for operators, there may be advantages to increasing the fidelity of the visualizations to help operators perform better. The improvements align with previous manual control research that suggests that VR can improve performance over desktop visualizations (Elor et al. 2021).

No differences were found in any of the SA levels. This rejects our hypothesis that VR or 3D visualizations will improve SA. Previous literature is mixed on the effects of display modality on SA.In a previous study for remote monitoring of satellites, both VR and 3D visualizations improved SA over traditional displays for level 2 and 3 SA, but that VR reduced level 1 SA (Buchner et al. 2025). In autonomous surface vehicle monitoring VR and 3D visualizations improved SA over traditional displays (Lager and Topp 2019), and 3D visualizations improved SA for air traffic control tasks (Rottermanner et al. 2020). These results indicate that although remote supervision relies heavily on monitoring, monitoring studies alone cannot fully capture the intricacies and changes in operation demands that supervision requires with regards to SA. Additionally, this may indicate that not all forms of operations may benefit from the additional immersion or visualization. These differences may be attributed to using different SA measurement techniques, and may be a limitation of the goal-directed task analysis to develop SPAM queries, as not all questions benefited from the use of visualizations, particularly for level 1 SA.

No differences were found in workload which supports our hypothesis and agrees with the previous satellite monitoring experiment. Due to the low workload nature of monitoring and supervision (Hooey et al. 2018; Huey and Wickens 1993; Grier 2015) with traditional displays it was unexpected that any one new display could significantly reduce workload overall. However, there is a concern that the novelty and physical aspects of the VR environment (walking around, using the controller) could have increased workload (Chao et al. 2017; Millais et al. 2018), but this was not found to be the case.

For usability, there were no differences in displays found which agrees with the previous work for satellite monitoring. However, this disagrees with other experiments that often subjectively have VR improve usability (Elor et al. 2021; Lager and Topp 2019). Notably these previous studies are within subjects. Using a between subject design is a limitation to our research. Participants are not able to compare to the other display modalities, but if they could compare and contrast the features and limitations of the three displays, stronger preferences might have been apparent.

The subjective utility found that 3D visualizations (either on a screen or in VR) can improve the participant's understanding of orbital motion. This is a key aspect for future satellite operations, as complex relative orbits can be difficult to understand intuitively (Sellmaier and Frei 2022). This also may explain some of the differences in performance. This conclusion is supported by the majority of the subjective comments about the Baseline display which included a dislike of the orbit paths and not being able to visualize them.

There are some limitations to our study. Our participants may not be representative of trained operators who would use these types of displays and had varying backgrounds in orbital mechanics and operations. This was taken into account by including demographics in the statistical models, but having a more representative subject pool would be beneficial. While this study included limited operator control over the satellite systems, the scenario was not as operationally complex as real world scenarios which may also include aspects such as limited communication, anomalies, more telemetry streams, and slower and prolonged operations. The short time span likely reduced the amount of boredom participants experienced and may have reduced the amount of discomfort the VR headset caused. VR, for prolonged use, may be uncomfortable due to the headset weight and eyestrain (Chen et al. 2021). While our study did not elicit these types of responses from subjects, it does limit the ability answer these research questions with increased experimental fidelity.

Future work involves expanding these experiments to more complex operations and using trained operators to understand the effects of VR. It should also include different use cases for VR within operations. While VR may not be promising for use in continuous operations, the subjective responses and utility show that VR may offer benefits as a way to intuitively understand the orbits. This may mean that it could be useful in advanced planning of operations to understand the orbits, visualizing the environment during a difficult or sensitive maneuver, or improving training.

The findings of this work has implications when designing remote supervisory spaceflight displays. Particularly, 3D visualizations may improve performance in complex situations and allow improved operator understanding of difficult-to-understand trajectories or motions. Therefore, it may be beneficial to design or modify displays to incorporate 3D visualizations. However, VR may not be beneficial to incorporate into remote supervisory displays in its current state due to the lack of further benefits. Additionally, incorporating 3D visualizations would require a lower barrier to entry as compared to VR, and it may be feasible to modify existing display setups and mission control rooms to include them. These implications may also extend to displays for other supervisory-based operations, such as air traffic control, maritime management, manufacturing, or exploration, especially those that present similar challenges to spaceflight operations. These types of paradigms are critical and important for aspects of daily life, and future work can extend this research into these fields to understand if the findings are true across applications or application-specific.

5 Conclusion

This study compares the effects of 3D visualization and immersion for remote supervision of a spacecraft operation on performance, SA, workload, usability, and utility. Three displays were designed for a satellite rendezvous task and compared via a human subject experiment. The results of this work indicate that 3D visualizations, whether on a computer screen or in VR, may improve utility and performance in hard trials. However, our results show that there is little evidence that immersion through VR provides additional benefits. There are no differences between displays in SA, workload, and usability. While VR has been shown in other research to provide benefits in manual control operations, it does not translate to our findings for remote supervision. Future work can assess other use cases for VR in operations, such as forms of transportation or manufacturing. These conclusions can also inform other supervisory operations that are emerging, such as transportation, manufacturing, and robotics.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10055-025-01263-5.

Acknowledgements We would like to thank Mikaela Felix and Jack Fox for their work in the development of the task, and everyone who participated in the experiment.

Author contributions S.B.: Conceptualization, Formal analysis, Methodology, Investigation, Visualization, Writing—original draft J.W.: Software, Writing—review & editing G.S.: Methodology, Investigation, Writing—review & editing B.B.: Investigation, Writing—review & editing H.S.: Funding acquisition, Supervision, Project Administration, Writing—review & editing AH: A.H.: Conceptualization, Funding acquisition, Methodology, Supervision, Project Administration, Writing—review & editing

Funding This work was funded by SpaceWerx grant AWD-23-04-0184.

Data availability Data is available upon request.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose. The authors declare no Conflict of interest.

Ethical approval The study was approved by the University of Colorado Institutional Review Board (#24-0250).

Informed consent Informed consent was obtained from all participants.

References

ARAC WG Report (2004) ARAC WG Report FAR/JAR 25.1322 & AC/ACJ 25.1322. Tech. rep., Federal Aviation Administration, https://www.faa.gov/regulations_policies/rulemaking/committees/documents/media/TAEas-wcal-04232002.pdf

Alexis D (2014) dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums. https://doi.org/10.32614/CRAN.package.dunn.test, https://CRAN.R-project.org/package=dunn.test, institution: Comprehensive R Archive Network Pages: 1.3.6

Basner M, Savitt A, Moore TM et al (2015) Development and validation of the cognition test battery for spaceflight. Aerosp Med Hum Perform 86(11):942–952. https://doi.org/10.3357/AMHP.4343.2015

Bates D, Mächler M, Bolker B et al (2015) Fitting linear mixed-effects models using lme4. J Stat Softw 67:1–48. https://doi.org/10.18637/jss.v067.i01

Ben-Shachar MS, Lüdecke D, Makowski D (2020) Effectsize: estimation of effect size indices and standardized parameters. J Open Sour Softw 5(56):2815. https://doi.org/10.21105/joss.02815

Brooke J (1996) SUS: a quick and dirty usability scale. In: Usability evaluation in industry. CRC Press, p 6

Bualat M, Fong T, Allan M, et al (2013) Surface telerobotics: development and testing of a crew controlled planetary rover system. In: AIAA SPACE 2013 Conference and Exposition. American Institute of Aeronautics and Astronautics, San Diego, CA, https://doi.org/10.2514/6.2013-5475

Buchner SL, Rindfuss A, Wood J, et al (2025) Impacts of 3D visualizations and virtual reality in display designs for remote monitoring of satellite operations. Frontiers in virtual Reality

Caroux L, Isbister K (2016) Influence of head-up displays' characteristics on user experience in video games. Int J Hum Comput Stud 87:65–79. https://doi.org/10.1016/j.ijhcs.2015.11.001 (https://www.sciencedirect.com/science/article/pii/S1071581915001779)

Chao CJ, Wu SY, Yau YJ et al (2017) Effects of three-dimensional virtual reality and traditional training methods on mental workload and training performance. Hum Factors Ergon Manuf Serv Ind 27(4):187–196. https://doi.org/10.1002/hfm.20702

Chen JYC, Haas EC, Barnes MJ (2007) Human performance issues and user interface design for teleoperated robots. IEEE Trans Syst Man Cybern Part C (Appl Rev) 37(6):1231–1245. https://doi.org/10.1109/TSMCC.2007.905819. (**conference Name: IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)**)

Chen Y, Wang X, Xu H (2021) Human factors/ergonomics evaluation for virtual reality headsets: a review. CCF Trans Pervas Comput Interact 3(2):99–111. https://doi.org/10.1007/s42486-021-00062-6

Christense RHB (2023) ordinal—Regression Models for Ordinal Data. http://cran.nexr.com/web/packages/ordinal/ordinal.pdf

Cordeil M, Dwyer T, Hurter C (2016) Immersive solutions for future Air Traffic Control and Management. In: Proceedings of the 2016 ACM Companion on Interactive Surfaces and Spaces. ACM, Niagara Falls Ontario Canada, pp 25–31, https://doi.org/10.1145/3009939.3009944

Cunningham JC, Battiste H, Curtis S et al (2015) Measuring situation awareness with probe questions: reasons for not answering the probes. Proc Manuf 3:2982–2989. https://doi.org/10.1016/j.promfg.2015.07.840 (https://www.sciencedirect.com/science/article/pii/S2351978915008410)

Dan A, Reiner M (2017) EEG-based cognitive load of processing events in 3D virtual worlds is lower than processing events in 2D displays. Int J Psychophysiol 122:75–84. https://doi.org/10.1016/j.ijpsycho.2016.08.013 (https://www.sciencedirect.com/science/article/pii/S0167876016306882)

Davis S, Nesbitt K, Nalivaiko E (2014) A systematic review of cybersickness. In: Proceedings of the 2014 conference on interactive entertainment. ACM, Newcastle NSW Australia, pp 1–9, https://doi.org/10.1145/2677758.2677780

Durso FT, Hackworth CA, Truitt TR, et al (1999) Situation awareness as a predictor of performance in en route air traffic controllers.: (402702004-001). Tech. Rep. DOT/FAA/AM-99/3, American Psychological Association, http://doi.apa.org/get-pe-doi.cfm?doi=10.1037/e402702004-001

Durso FT, Hackworth CA, Truitt TR et al (1998) Situation awareness as a predictor of performance for en route air traffic controllers. Air Traffic Control Quart 6(1):1–20. https://doi.org/10.2514/atcq.6.1.1. (**publisher: American Institute of Aeronautics and Astronautics**)

Elor A, Thang T, Hughes BP, et al (2021) Catching jellies in immersive virtual reality: a comparative teleoperation study of Rovs in underwater capture tasks. In: Proceedings of the 27th ACM symposium on virtual reality software and technology. ACM, Osaka Japan, pp 1–10, https://doi.org/10.1145/3489849.3489861

Endsley MR (1988) Design and evaluation for situation awareness enhancement. Proc Hum Fact Soc Annu Meet 32(2):97–101. https://doi.org/10.1177/154193128803200221. (**publisher: SAGE Publications**)

Endsley MR (2000) Direct measurement of situation awareness: Validity and use of SAGAT. Situation awareness analysis and measurement. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, pp 147–173

Fujino M, Lee J, Hirano T, et al (2020) Comparison of SAGAT and SPAM for seeking effective way to evaluate situation awareness and workload during air traffic control task. In: Proceedings of the 2020 HFES 64th international annual meeting, virtual, https://doi.org/10.1177/1071181320641442

Gad R, Martin S, Olbrich M, et al (2023) Towards Leveraging Augmented and Virtual Reality for Spacecraft Mission Operations at ESOC. In: 17th International Conference on Space Operations, Dubai, UAE

Golding JF (1998) Motion sickness susceptibility questionnaire revised and its relationship to other forms of sickness. Brain Res Bull 47(5):507–516. https://doi.org/10.1016/s0361-9230(98)00091-4

Gorbunov AL, Nechaev EE (2022) Augmented reality technologies in air transport control systems. In: 2022 systems of signals generating and processing in the field of on board communications, pp 1–5, https://doi.org/10.1109/IEEECONF53456.2022.9744399

Grier R (2015) How high is high? A metanalysis of NASA TLX global workload scores. https://doi.org/10.1177/1541931215591373

Hart SG, Staveland LE (1988) Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Hancock PA, Meshkati N (eds) Advances in psychology, human mental workload, vol 52. North-Holland, pp 139–183, https://doi.org/10.1016/S0166-4115(08)62386-9

Hooey BL, Kaber DB, Adams JA et al (2018) The underpinnings of workload in unmanned vehicle systems. IEEE Trans Hum-Mach Syst 48(5):452–467. https://doi.org/10.1109/THMS.2017.2759758. (**conference Name: IEEE Transactions on Human-Machine Systems**)

Hosseini A, Lienkamp M (2016) Enhancing telepresence during the teleoperation of road vehicles using HMD-based mixed reality.

In: 2016 IEEE intelligent vehicles symposium (IV), pp 1366–1373, https://doi.org/10.1109/IVS.2016.7535568

Huey BM, Wickens CD (1993) Workload Transition: Implications for Individual and Team Performance. Tech. rep, National Research Council

Iacovides I, Cox A, Kennedy R, et al (2015) Removing the HUD: the impact of non-diegetic game elements and expertise on player involvement. In: Proceedings of the 2015 annual symposium on computer-human interaction in play. Association for computing machinery, New York, NY, USA, CHI PLAY '15, pp 13–22, https://doi.org/10.1145/2793107.2793120

Kalamkar S, Biener V, Beck F, et al (2023) Remote monitoring and teleoperation of autonomous vehicles–is virtual reality an option? In: 2023 IEEE international symposium on mixed and augmented reality (ISMAR), pp 463–472, https://doi.org/10.1109/ISMAR59233.2023.00061

Kenneally PW, Piggott S, Schaub H (2020) Basilisk: a flexible, scalable and modular astrodynamics simulation framework. J Aerosp Inf Syst 17(9):496–507. https://doi.org/10.2514/1.I010762. (**publisher: American Institute of Aeronautics and Astronautics**)

Kuznetsova A, Brockhoff PB, Christensen RHB (2017) lmerTest package: tests in linear mixed effects models. J Stat Softw 82:1–26. https://doi.org/10.18637/jss.v082.i13

Lager M, Topp EA (2019) Remote supervision of an autonomous surface vehicle using virtual reality. IFAC-PapersOnLine 52(8):387–392. https://doi.org/10.1016/j.ifacol.2019.08.104 (https://linkinghub.elsevier.com/retrieve/pii/S2405896319304380)

Lejuez CW, Read JP, Kahler CW et al (2002) Evaluation of a behavioral measure of risk taking: the balloon analogue risk task (BART). J Exp Psychol Appl 8(2):75–84. https://doi.org/10.1037/1076-898x.8.2.75

Lenth RV, Bolker B, Buerkner P, et al (2023) emmeans: Estimated Marginal Means, aka Least-Squares Means. https://cran.r-project.org/web/packages/emmeans/index.html

Loft S, Morrell D, Huf S (2013) Using the situation present assessment method to measure situation awareness in simulated submarine track management. Int J Hum Factors Ergon 2:33–48. https://doi.org/10.1504/IJHFE.2013.055975

Loft S, Bowden V, Braithwaite J et al (2015) Situation awareness measures for simulated submarine track management. Hum Factors 57(2):298–310. https://doi.org/10.1177/0018720814545515

Marre Q, Caroux L, Sakdavong JC (2021) Video game interfaces and diegesis: the impact on experts and novices' performance and experience in virtual reality. Int J Hum-Comput Interact 37(12):1089–1103. https://doi.org/10.1080/10447318.2020.1870819

Maxime HERVE (2011) RVAideMemoire: Testing and Plotting Procedures for Biostatistics. https://doi.org/10.32614/CRAN.package.RVAideMemoire

mckenzie c, Glazier A (2017) Designing screen interfaces for vr. https://www.youtube.com/watch?v=ES9jArHRFHQ

Millais P, Jones SL, Kelly R (2018) Exploring data in virtual reality: comparisons with 2D data visualizations. In: Extended Abstracts of the 2018 CHI conference on human factors in computing systems. ACM, Montreal QC Canada, pp 1–6, https://doi.org/10.1145/3170427.3188537

Mirchi T, Vu KP, Miles J et al (2015) Air traffic controller trust in automation in NextGen. Proc Manuf 3:2482–2488. https://doi.org/10.1016/j.promfg.2015.07.509

O'Hara J, Fleger S (2020) Human-System Interface Design Review Guidelines. Tech. Rep. BNL–216211-2020-FORE, 1644018, https://doi.org/10.2172/1644018

Read JM, Saleem JJ (2017) Task performance and situation awareness with a virtual reality head-mounted display. Proc Hum Factors Ergon Soc Annu Meet 61(1):2105–2109. https://doi.org/10.1177/1541931213602008

Reed SC, Levin FR, Evans SM (2012) Alcohol increases impulsivity and abuse liability in heavy drinking women. Exp Clin Psychopharmacol 20(6):454–465. https://doi.org/10.1037/a0029087. (**place: US Publisher: American Psychological Association**)

Rosyid HA, Pangestu AY, Akbar MI (2021) Can diegetic user interface improves immersion in role-playing games? In: 2021 7th international conference on electrical, electronics and information engineering (ICEEIE), pp 200–204, https://doi.org/10.1109/ICEEIE52663.2021.9616732

Rottermanner G, de Jesus Oliveira VA, Lechner P, et al (2020) Design and evaluation of a tool to support air traffic control with 2D and 3D visualizations. In: 2020 IEEE conference on virtual reality and 3D user interfaces (VR), pp 885–892, https://doi.org/10.1109/VR46266.2020.00011

Sellmaier F, Frei H (2022) Operations of on-orbit servicing missions. In: Sellmaier F, Uhlig T, Schmidhuber M (eds) Spacecraft operations. Springer International Publishing, Cham, pp 491–529, https://doi.org/10.1007/978-3-030-88593-9_25

Sheridan TB (2012) Human supervisory control of automation. In: Handbook of human factors and ergonomics. John Wiley & Sons, Incorporated, New York, http://ebookcentral.proquest.com/lib/ucb/detail.action?docID=817338

Sheridan TB (1992) Musings on telepresence and virtual presence. Presence Teleop Virt 1(1):120–126. https://doi.org/10.1162/pres.1992.1.1.120

Sim L, Cummings ML, Smith CA (2008) Past, present and future implications of human supervisory control in space missions. Acta Astronaut 62(10):648–655. https://doi.org/10.1016/j.actaastro.2008.01.029 (https://www.sciencedirect.com/science/article/pii/S0094576508001045)

Sittner F, Hartmann O, Montenegro S, et al (2023) An update on the virtual mission control room. In: 37th annual small satellite conference, Logan Utah, https://doi.org/SSC23-P5-25

Slater M, Linakis V, Usoh M, et al (1996) Immersion, presence and performance in virtual environments: an experiment with tri-dimensional chess. In: Proceedings of the ACM symposium on virtual reality software and technology - VRST '96. ACM Press, Hong Kong, pp 163–172, https://doi.org/10.1145/3304181.3304216

Tittle JS, Roesler A, Woods DD (2002) The remote perception problem. Proc Hum Factors Ergon Soc Annu Meet 46(3):260–264. https://doi.org/10.1177/154193120204600309. (**publisher: SAGE Publications Inc**)

Tsigkounis K, Komninos A, Politis N, et al (2021) Monitoring maritime industry 4.0 systems through VR environments. In: CHI Greece 2021: 1st international conference of the ACM Greek SIGCHI Chapter. ACM, Online (Athens, Greece) Greece, pp 1–8, https://doi.org/10.1145/3489410.3489429

van Emmerik ML, de Vries SC, Bos JE (2011) Internal and external fields of view affect cybersickness. Displays 32(4):169–174. https://doi.org/10.1016/j.displa.2010.11.003 (https://www.sciencedirect.com/science/article/pii/S0141938210000934)

Van de Merwe DB, Van Maanen L, Ter Haar FB, et al (2019) Human-robot interaction during virtual reality mediated teleoperation: how environment information affects spatial task performance and operator situation awareness. In: Chen JY, Fragomeni G (eds) Virtual, augmented and mixed reality. Applications and case studies. Springer International Publishing, Cham, Lecture Notes in Computer Science, pp 163–177, https://doi.org/10.1007/978-3-030-21565-1_11

Warm S, Matthews G, Finomore V (2008) Vigilance, workload, and stress. Performance Under Stress pp 115–141

Whitney D, Rosen E, Phillips E, et al (2020) Comparing Robot Grasping Teleoperation across desktop and virtual reality with ROS reality. In: Amato NM, Hager G, Thomas S, et al (eds) Robotics research, Springer Proceedings in Advanced Robotics. vol 10.

Springer International Publishing, Cham, pp 335–350, https://doi.org/10.1007/978-3-030-28619-4_28

Whitney D, Rosen E, Phillips E, et al (2020) Comparing robot grasping teleoperation across desktop and virtual reality with ROS reality. In: Amato NM, Hager G, Thomas S, et al (eds) Robotics research. Springer International Publishing, Cham, Springer Proceedings in Advanced Robotics, pp 335–350, https://doi.org/10.1007/978-3-030-28619-4_28

Wickens CD (1993) Cognitive factors in display design. J Washington Acad Sci 83(4):179–201

Wood J, Margenet MC, Kenneally P, et al (2018) Flexible basilisk astrodynamics visualization software using the unity rendering engine. In: AAS Guidance and Control Conference, Breckenridge, CO

Woods NBSDavid D. (2011) Situation awareness: a critical but ill-defined phenomenon. In: Situational awareness. Routledge, p 14

Yao R, Heath T, Davies A et al (2014) Oculus VR best practices guide. Tech. rep, Oculus