An Evaluation of Visual and Performance Metrics in Virtual Reality-Based Parachute Descent Training

Jenna Korentsides, Victoria Trabysh
Embry-Riddle Aeronautical University
Daytona Beach, Florida
korentsj@my.erau.edu, trabyshv@my.erau.edu

Emily Rickel, Matthew B. Pierce
Naval Air Warfare Center Training Systems Division
Orlando, Florida
emily.a.rickel.civ@us.navy.mil, matthew.b.pierce7.civ@us.navy.mil

Barbara Chaparro, Joseph R. Keebler
Embry-Riddle Aeronautical University
Daytona Beach, Florida
chaparb1@erau.edu, keeblerj@erau.edu

Beth F. Wheeler Atkinson
Naval Air Warfare Center Training Systems Division
Orlando, Florida
beth.f.atkinson.civ@us.navy.mil

ABSTRACT

Virtual Reality (VR) has emerged as a promising training tool across various domains due to its immersive and cost-effective nature. However, its effectiveness is influenced by factors such as technical limitations, adverse side effects, and individual differences. This study explores the effectiveness of VR-based parachute descent training, focusing on the impact of Field of Regard (FOR) on task performance, seat kit release accuracy, course correction, and trainee confidence. Thirty-five participants engaged in VR parachute descent training sessions using an HTC Vive Pro headset. The analysis employed Bayesian methods to evaluate the effects of FOR variations on the key performance indicators. Results from Bayesian analyses suggest that FOR variations did not significantly affect task performance, seat kit release accuracy, course correction, confidence levels, time to first fixation, dwell time, total number of fixations, or time spent searching. Specifically, the posterior distributions for differences between groups with varying FORs showed a high probability of being close to zero for all measured variables, indicating no substantial impact of FOR on these aspects of VR training performance. These findings challenge the assumption that larger FOR enhances VR training effectiveness. It suggests that other factors (e.g., content quality, instructional design, and individual differences), may play more critical roles in determining the success of VR-based training programs. Overall, these findings provide valuable insights into the optimal design of VR training environments, suggesting that larger FOR may not always result in improved training outcomes. Additionally, considerations for training implications and cost-effectiveness are discussed, emphasizing the importance of evaluating technology before adoption to ensure desired training outcomes are achieved while minimizing unnecessary expenses. This research contributes to the understanding of VR training effectiveness and informs future developments in VR-based training methodologies.

ABOUT THE AUTHORS

Jenna Korentsides is a PhD in Human Factors candidate at Embry-Riddle Aeronautical University (ERAU). She works as a graduate research assistant in the Small Teams Analog Research (STAR) lab. Jenna holds a B.A. in psychology from Stockton University and an M.S. in Human Factors from ERAU. Presently, Jenna continues as a PhD candidate with her primary research expertise including human-computer interaction, human-agent/human-AI interaction/teaming, training, teamwork, statistical analysis and modeling, and user experience.

Victoria Trabysh is a Research Psychologist for the DoD/Air Force. She works under headquarters Air Education and Training Command, performing test and evaluation on aircraft systems and military pilot training. She contributed to the work featured in this paper while she was a student at ERAU, where she earned her M.S. in Human Factors.

Her primary research interests are aviation psychology, training science, and human performance. Victoria also holds a bachelor's degree in psychology from Trinity University.

Emily Rickel is a Research Psychologist at the Naval Air Warfare Center Training Systems Division (NAWCTSD) in the Basic and Applied Training and Technology for Learning and Evaluation (BATTLE) Laboratory. Her research interests include extended reality training and user interface design/analysis. Dr. Rickel earned a Ph.D. in Human Factors from ERAU.

Matthew B. Pierce is a Research Psychologist at NAWCTSD in the BATTLE Laboratory. His master's degree is in Human Factors from ERAU where he conducted research to explore differences in virtual reality (VR) and screen-based performance. His research interests include biometrics, human factors analyses of training platforms, and incorporation of modern technology to advance aviation training sciences.

Barbara Chaparro has a PhD in Experimental Psychology from Texas Tech University. She is a Professor in the Human Factors and Behavioral Neurobiology Department and is head of the Research in User eXperience (RUX) Lab at ERAU. Her research interests include the study of factors that influence the user experience (UX) of products, software and systems, the investigation of usability assessment methods, and the efficacy of augmented/mixed reality devices and applications.

Joseph R. Keebler has over 15 years of experience conducting experimental and applied research in human factors, with a specific focus on training and teamwork in medical, military, and consumer domains. He has partnered with multiple agencies and has led projects aimed at the implementation of HF/E in complex, high-risk systems to increase safety and human performance. This work includes command and control of tele-operated unmanned systems, communication and teamwork in medical systems, and simulation-/game- based training for advanced skills. His work includes over 50 publications and over 60 presentations at national and international conferences. He is currently director of the Small Teams Analog Research (STAR) Laboratory and is co-director of the Research Engineering and Applied Collaborations in Healthcare (REACH) Laboratory at ERAU where he works with a team of faculty and students to solve theoretical and practical teamwork and medical issues.

Beth F. Wheeler Atkinson is a Senior Research Psychologist at NAWCTSD, a Naval Air Warfare Center Aircraft Division (NAWCAD) Fellow, and lead of the BATTLE Laboratory. She has led and transitioned several research and development efforts devoted to investigating capability enhancements for training and operational environments, with research interests in instructional technologies, Human Computer Interaction/user interface design and analysis, and aviation safety training. Ms. Atkinson earned an M.A. in Psychology, Applied Experimental Concentration, from the University of West Florida.

An Evaluation of Visual and Performance Metrics in Virtual Reality-Based Parachute Descent Training

Jenna Korentsides, Victoria Trabysh
Embry-Riddle Aeronautical University
Daytona Beach, Florida
korentsj@my.erau.edu, trabyshv@my.erau.edu

Emily Rickel, Matthew B. Pierce
Naval Air Warfare Center Training Systems Division
Orlando, Florida
emily.a.rickel.civ@us.navy.mil, matthew.b.pierce7.civ@us.navy.mil

Barbara Chaparro, Joseph R. Keebler
Embry-Riddle Aeronautical University
Daytona Beach, Florida
chaparb1@erau.edu, keeblerj@erau.edu

Beth F. Wheeler Atkinson
Naval Air Warfare Center Training Systems Division
Orlando, Florida
beth.f.atkinson.civ@us.navy.mil

INTRODUCTION

Parachute Descent Training

Parachute descent training is a critical aspect of aviation survival education that prepares naval personnel for airborne operations, ensuring they possess the necessary skills to execute safe and effective parachute procedures in emergency scenarios. This training is conducted at an aviation survival training center (ASTC) using a combination of lectures and a simulation-based trainer for practical exercises. Trainees receive classroom instruction covering topics such as parachute equipment, emergency procedures, aerodynamics, and landing techniques. This foundation emphasizes the principles behind parachute descent and methods or techniques that enhance safety. Following classroom instruction, trainees proceed to practical training in a parachute descent trainer (PDT) that simulates real-world scenarios to support practicing skills such as malfunction identification and mitigation, decision-making for landing locations, canopy control, and body position for safe landings. This hands-on training enables familiarization with the equipment and develops an understanding of procedures that affect survivability.

To enhance PDTs, the Navy is exploring capabilities that support ease of use and facilitate instructor-student discussions related to trainee performance and decision-making. Additional challenges to training throughput under consideration include system performance (e.g., reliability, calibration requirements), sense of immersion, capture of student’s actions, and instructor aids to guide performance feedback. The technology under development, SkyFall, uses a contemporary virtual environment to provide realistic visuals of the environment, weather conditions, and the parachute’s descent. The system design is flexible, providing options for either a display-based virtual environment or virtual reality (VR) head-mounted display (HMD), as well as a means to monitor student actions through system-based event markers and instructor observations.

The primary emphasis of parachute descent training within the survival training environment is a comprehensive set of procedures and consideration of decision-making factors essential for aviators' safety and survival during emergency situations. Specifically, instructional content includes how to execute emergency procedures in the event of parachute malfunctions. This involves practicing corrective actions for canopy malfunctions to build skills to troubleshoot and resolve issues swiftly and effectively. The cornerstone of this training is memorization of IROK – an acronym that stands for inflate, release, options, and Koch fittings – which is intended to aid aviators during training to successfully perform proper procedures and decision-making. The first step, inflate, refers to inflating of an individual's life preserver unit (LPU) to facilitate floatation in the event of water landing or to provide additional cushion for neck protection. The second step is the release of the aviator's seat kit. Releasing the seat kit is necessary to prevent injury but also requires consideration of timing as the emergency survival contents contained in the kit must be located after landing to support survival and rescue. The options aspect of IROK focuses on decision-making that is critical to the safety and survivability of the landing. Factors to consider include equipment to shed or leave on to offer increased protection (e.g., visor) and landing area options and ability to steer toward those locations to reduce hazardous conditions. Finally, Koch fittings refers to unclasping the quick release fittings that connect the aviator's harness to the parachute (this step was not examined in the current study).

Throughout practical exercise training, instructors observe students to evaluate performance related to IROK procedures. An objective, rubric-based assessment ensures that trainees understand the necessary procedures and consider critical decision-making factors to help prepare them for the demands of potential real-world emergencies to enhance readiness and survivability.

Virtual Reality Training

VR fully immerses users in an entirely simulated digital environment. This immersion can be achieved through the utilization of HMDs that occlude the user's view of the real world (Stanney et al., 2021). VR technologies have been leveraged as training tools across several domains, including healthcare, maintenance, aviation, and manufacturing (Stanney et al., 2021; Xie et al., 2021). VR has several advantageous traits that make it an appealing training tool. For instance, VR training is more cost-effective, immersive, flexible, and safer compared to training in field environments (Kaplan et al., 2021; Moroney & Lilienthal, 2009). However, there are also multiple factors that can negatively impact the utilization of VR as an effective training solution, including technical limitations (e.g., battery capacity, bulky hardware, latency issues), adverse side effects experienced by the user (e.g., cybersickness), and individual differences (e.g., trainee learning styles, acceptance of VR technology, spatial abilities; Kaplan et al., 2021; Peracchio, 2020; Stanney et al., 2021).

Research evaluating the effectiveness of VR-based training has expanded in recent years (Strojny & Duzmanska-Misiarczyk, 2023), likely due to the rising adoption of VR training technologies as a result of decreased cost and increased capabilities (Stanney et al., 2021). VR training effectiveness is often measured using knowledge and skill tests, as well as subjective measures typically collected through questionnaires and interviews that provide insight into user perceptions of the training experience, such as usability, self-efficacy, and engagement. Observation by an expert and physiological measurements (e.g., eye-tracking, galvanic skin response, heart rate variability) have also been used to evaluate VR training effectiveness (Strojny & Duzmanska-Misiarczyk, 2023; Xie et al., 2021).

Multiple review articles that summarize several individual investigations of VR training effectiveness have been recently published. These findings exhibit a range of conclusions, with some reviews suggesting that VR training is superior compared to alternative training methods (Abich et al., 2021; Angel-Urdinola et al., 2021; Howard et al., 2021; Radhakrishnan et al., 2021). Other reviews have found that VR and alternative training modalities are comparable (Kaplan et al., 2021), or were inconclusive due to variability between the findings of included studies (Hepperle & Wolfel, 2023; Narciso et al., 2021). Overall, mixed results across these review articles may be attributed to disparities in task-technology fit (Abich et al., 2021; Howard et al., 2021), population (Kaplan et al., 2021), and scientific rigor (Strojny & Duzmanska-Misiarczyk, 2023).

Field of View and Field of Regard

A user's experience within a simulated environment, including those displayed through VR HMDs, can be affected by field of view (FOV) and field of regard (FOR). FOV refers to the extent of the visual field that can be viewed instantaneously, while FOR refers to the total size of the visual environment encompassing the user (see Figure 1). Both FOV and FOR are measured in degrees of visual angle (Bowman & McMahan, 2007). Studies that investigated FOR in projection-based VR environments (i.e., Cave Automatic Virtual Environments, CAVEs) found that larger FOR was associated with fewer errors (Ragan et al., 2013), as well as faster searching behaviors and increased perceived usability (McMahan, 2011). In a study that compared a CAVE system to a VR HMD, Clifford et al. (2020) found that the larger FOR afforded by the HMD invoked greater situational awareness and presence, and was associated with decreased workload.

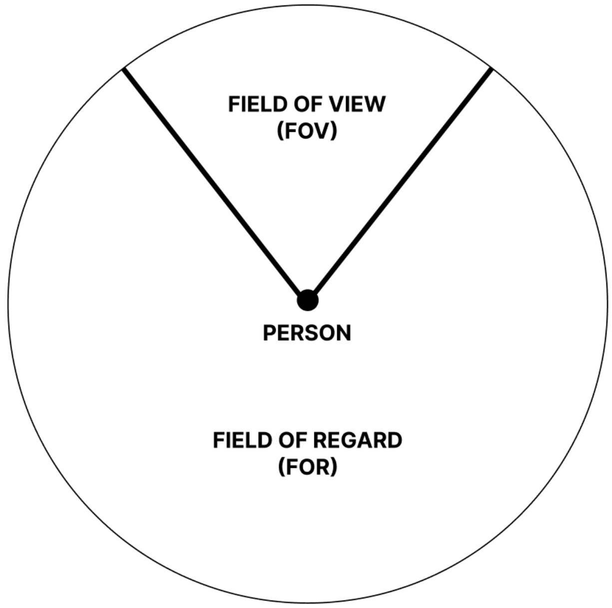

Figure 1. Representation of FOV and FOR

METHOD

Participants

The sample included 35 participants (23 males; 12 females). Of these participants, the majority (62.9%) reported that they have either never used or rarely use VR products (e.g., HTC Vive, HP, Oculus/Meta, Varjo), 31.5% reported using VR occasionally or a moderate amount, and only 5.7% reported being regular users of VR. Hours per week spent playing video games varied across the sample with 17.1% reporting 0 hours, 57.1% reporting 1-3 hours, 11.4% reporting 4-6 hours, 5.7% reporting 7-9 hours, and 8.6% reporting 10+ hours. None of the participants had military parachute jumping experience; however, 20% of participants had experienced civilian parachute jumping at least one time (17.1% of these reported 1 time; 2.9% reported 2-5 times).

Two participants' data was excluded from the analysis due to data integrity loss. One individual requested to deflate the LPU that participants were required to wear while seated in the simulator (for comfort purposes), while another failed to follow directions provided by the facilitating researcher. Individuals were not eligible to participate in this study if they had a history of seizures and/or any back injuries given the visual stimulus within the virtual environment and the physical movement required. An informed consent document outlining these parameters was signed by each participant prior to beginning the study.

Materials

The parachute descent training simulator incorporates a range of materials and features to provide a comprehensive and immersive training experience for naval personnel. The simulator features realistic 3D visual environments depicting various terrain and weather conditions and user controls that enable trainees to interact with the virtual environment to practice emergency procedures and to refine skills in a safe and controlled setting. Users sat in a tethered bucket seat attached to the steel frame of the simulator as shown in Figure 2. Two yellow steering toggles hung down on either side of the seat to control steering throughout the simulated descent. Additionally, an LPU-like vest (i.e., a life vest that ASTC subject-matter experts confirmed was a close match to what pilots use) was inflated around participants necks to simulate head movement restrictions that would result after completion of the Inflate step of IROK.

Figure 2. Image of Simulator

Virtual Reality Head-Mounted Display

The simulated training environment was presented to participants through the HTC Vive Pro, a VR HMD released in 2018 (Vive, 2018). The HTC Vive Pro features a 110-degree FOV displayed through a 3.5-inch diagonal screen (Vive, n.d.).

Measures

Spatial Orientation

Defined as the ability to perceive and maintain one's location relative to objects in space (American Psychological Association, 2018), spatial orientation was measured using an online version of the Guilford-Zimmerman Spatial Orientation Survey (1948). Participants were allotted 10 minutes to complete 59 questions. Each question displayed a diagram of a boat's starting position in the water with environmental cues, such as trees, other boats, and the horizon, from the point of view of sitting in the boat and looking over the bow. A second diagram represented the boat's current position after moving within the environment. Participants responded to each question by indicating how the boat moved between the two diagrams. Higher scores indicate better spatial orientation ability (Kyritsis & Gulliver, 2009).

Task Performance

Performance was measured using a scoring rubric developed in accordance with training objectives and performance checklists utilized by parachute descent instructors to assess trainee knowledge retention and behavior in the simulator. Rubric items included steering system activation, turning efficiency, and steering toward a safe landing zone. Researchers rated the extent to which participants satisfied each rubric item on a four-point scale (Needs Improvement, Developing, Sufficient, Above Average). Researchers also noted course correction altitude and seat-kit release altitude. Target ranges for these metrics were 1000-3500 feet for course correction altitude and 150-400 feet for seat-kit release altitude.

Eye-Tracking

The following eye-tracking metrics were collected through the VR headset during all three FOR conditions: 1) time to first fixation, 2) dwell time, 3) number of fixations, and 4) search time. Time to first fixation quantifies the amount of time it takes for a participant to look at a specific area of interest from the time of stimulus onset. Dwell time quantifies the amount of time a participant spends looking at a specific area of interest. Number of fixations quantifies the number of times a participant looks at, or fixates on, a specific area of interest. Finally, search time quantifies the amount of time it takes for a participant to look at a specific area of interest.

Self-Efficacy

Self-efficacy was measured using seven items developed by Hartley et al. (2022) to gauge participant perception of their ability to successfully complete parachute descent tasks. Items were rated on a seven-point Likert scale (1 – Strongly Disagree; 7 – Strongly Agree). Self-efficacy was measured three times throughout the study, following exposure to each FOR condition. Items were summed to produce overall self-efficacy scores that ranged from 7 (low self-efficacy) to 49 (high self-efficacy).

Simulator Sickness

Simulator sickness was measured using the Simulator Sickness Questionnaire (SSQ; Kennedy et al., 1993). The SSQ consists of sixteen symptoms grouped into three subscales: nausea, oculomotor, and disorientation. The severity of each symptom is rated on a four-point scale (None, Slight, Moderate, Severe). Participants completed the SSQ three times throughout the study, following exposure to each FOR condition.

Procedure

Upon arrival, participants were seated in a cubicle where they completed the demographics questionnaire and the spatial orientation survey. The latter required participants to view a series of two images and indicate the direction(s) in which an object had moved in relation to its surroundings. The purpose of this survey was to capture a baseline measure of each participant's sense of spatial orientation. Next, participants were instructed to follow along with a 10-minute presentation educating them on Emergency Parachute Descent Procedures. This presentation covered the basics of the tasks they were required to complete in the simulator.

Once the preliminary training was complete, participants were then seated in the parachute simulator where the facilitating researcher guided them through a walkthrough scenario and at least three practice scenarios to gain familiarity with the system. Tasks participants were expected to complete included: activate steering by pulling down on the steering toggles, steer left and right, identify wind direction, position oneself into the wind to slow descent, and drop seat kit when between 150 and 400 feet above ground. If proficiency in these tasks was achieved by the third practice scenario, participants moved onto the test scenarios; if proficiency was not achieved, up to two additional scenarios were administered. The conditions of the practice scenarios were similar, but not identical to those of the test scenarios. In the practice scenario, a higher drop-altitude, higher winds, and orientations not seen in the test runs were used. Participants did not wear the VR HMD for the practice runs and instead viewed the virtual environment on the two television monitors shown in Figure 2. The purpose of these differences was to familiarize the participants with the functionality of the system, but not expose them to the exact conditions of the test scenarios. This allowed for a more accurate evaluation of the system as a trainer rather than the ability to replicate what was done during practice.

There were nine different test scenarios that were counterbalanced among all participants; each participant completed three of the nine scenarios, in varying order. The location (North Atlantic), drop-altitude (3,500 ft.), time of day (noon), wind (moderate), and rain (no) conditions were kept constant among all scenarios, with the FOR (60°, 180°, 360°) and drop-orientation relative to the smoke coming from the crashed aircraft (0°, 90°, 270°) being the only two manipulated variables in the scenarios. The observing researcher utilized a grading rubric to record participants' performance on each of the required tasks.

Following each of the three test scenarios, participants completed two self-report surveys – the SSQ and a seven-item self-efficacy survey regarding confidence level in performing various tasks in the SkyFall VR system. The purpose of the SSQ was to closely monitor any symptoms experienced by participants and to ensure simulator sickness was not a confounding variable impacting performance. There were no significant reports of simulator sickness throughout the duration of the study. The self-efficacy survey allowed for assessment of participant confidence in performing the required tasks, as well as relationships between self-efficacy and actual performance.

Hypotheses

A prior front-end analysis evaluated a number of commercially available and emerging training solutions as well as a variety of non-technology-mediated instruction options (Hartley et al., 2022). During front-end analysis, instructor feedback included a request to consider a larger FOR to obtain a "more realistic" training environment. This paper covers an eye tracking study that was conducted to evaluate students' performance at various FORs when going through simulator-based parachute training using SkyFall. Students are expected to rely on visual cues while scanning the environment during parachute descent training procedures, but it is unclear what students visually focus on compared to what they learn to look at in classroom settings. This effort would explore FOR needs and assess the virtual environment to ensure SkyFall produces realistic training environments while also making recommendations of improvement to parachute descent training. Data was gathered at Embry-Riddle Aeronautical University with the current version of the SkyFall simulator by running university students.

Hypothesis 1: Task Performance – Smaller FORs will yield better performance as measured by the observer rubric indicating safe survival, injury, or failure for IROK Procedures and Steering training objectives. (360° < 180° < 60°).

Hypothesis 2: Seat Kit Release Accuracy – Larger FORs, will yield more accurate seat kit release performance compared to training performance (360° > 180° > 60°).

Hypothesis 3: Course Correction – Larger FORs will yield longer course correction durations after identifying a wind-cue (360° > 180° > 60°).

Hypothesis 4: Confidence in Training – Larger FORs will yield higher levels of confidence in performance (360° > 180° > 60°).

Hypothesis 5: Time to First Fixation – Smaller FORs will yield longer times to first fixation for predefined AOIs (360° > 180° > 60°).

Hypothesis 6: Dwell Time – Larger FORs will yield longer dwell times for predefined AOIs (360° < 180° < 60°).

Hypothesis 7: Total Number of Fixations – Larger FORs will yield smaller number of fixations per AOI (360° < 180° < 60°).

Hypothesis 8: Time Searching – Larger FORs will yield more total time spent searching to locate task-specific AOIs (360° > 180° > 60°).

RESULTS

In this study, we employed both Bayesian analyses and repeated measures analysis of variance (ANOVAs), respectively, to examine the impact of varying FORs on performance and perceptual measures while using a VR parachute descent simulation. Bayesian analyses allow for a more nuanced interpretation of data, providing a measure of evidence for or against our hypotheses, while ANOVAs enable the examination of mean differences across conditions. These dual approaches were selected to robustly test our hypotheses under the assumptions that VR training effectiveness can be influenced by the FOR.

Bayesian Analysis Overview

The Bayesian analysis was structured to assess the probability of the null hypothesis (i.e., no difference due to FOR variations) relative to our experimental hypothesis for each measured outcome. In the reported results, several statistical values, including Bayes Factor, Mauchly's Test of Sphericity, and significance levels, are used to interpret the data obtained from our Bayesian analysis. For a description of each of these values (see Table 1).

Table 1. Definitions of Statistical Values

Value Name	Description
Bayes Factor	A Bayes Factor less than 1 indicates support for the null hypothesis, suggesting that the data are more likely under the null model than the alternative. A Bayes Factor from 1-3 indicates anecdotal evidence, 3-10 suggests moderate evidence, and 10-30 provides strong evidence.
Mauchly's Test of Sphericity	Checks whether the variances of the differences between all combinations of related groups (FORs in our study) are equal. If sphericity is assumed, it means the variances are equal, which is an assumption required for some statistical tests, including repeated measures

	ANOVA. A significant result (p < .05) indicates a violation of this assumption, which can affect the validity of the ANOVA results.
Significant Levels	The significance level (p-value) indicates the probability of observing the test statistic, given that the null hypothesis is true. A lower p-value (typically less than .05) suggests that it is unlikely to observe such data by chance alone, indicating a potential rejection of the null hypothesis.

Bayesian Analysis Results

Hypothesis 1 (Task Performance)
Task performance was expected to vary with FOR size, predicting superior results in smaller FORs. However, the Bayesian analysis indicated a Bayes Factor of .014, suggesting inadequate evidence to reject the null hypothesis. Similarly, Mauchly's Test of Sphericity was significant, W = 7.28, p = .004, indicating potential violations of the sphericity assumption that were not supported by the data.

Hypothesis 2 (Seat Kit Release Accuracy)
Predicting more accurate performance with larger FORs, the data yielded a Bayes Factor of .013, showing insufficient evidence against the null hypothesis. Mauchly's Test of Sphericity was non-significant, W = .975, p = .642.

Hypothesis 3 (Course Correction Time)
Larger FORs were hypothesized to increase the course correction times after identifying a wind cue. The Bayesian analysis yielded a Bayes Factor of .036, indicating insufficient evidence against the null hypothesis, and Mauchly's Test of Sphericity was non-significant, W = .956, p = .455.

Hypothesis 4 (Confidence in Training)
We expected that larger FORs would improve confidence in performance. The Bayesian analysis reported a Bayes Factor of .270, insufficient to reject the null hypothesis, while Mauchly's Test of Sphericity indicated significant sphericity, W = .509, p < .001.

Hypothesis 5 (Time to First Fixation)
It was predicted that smaller FORs would result in longer times to first fixation. The Bayes Factor was .054, providing insufficient evidence to support the hypothesis. Mauchly's Test of Sphericity was non-significant, W = .889, p = .135.

Hypothesis 6 (Dwell Time)
The hypothesis that larger FORs would yield longer dwell times was tested. The Bayesian analysis resulted in a Bayes Factor of .012, indicating insufficient evidence against the null hypothesis. Mauchly's Test of Sphericity was also non-significant, W = .937, p = .331.

Hypothesis 7 (Number of Fixations)
Predictions were that larger FORs would yield a smaller number of fixations per area of interest (AOI). The Bayesian analysis gave a Bayes Factor of .045, insufficient to support the hypothesis, and Mauchly's Test of Sphericity was significant, W = .730, p = .005.

Hypothesis 8 (Time Searching)
This hypothesis stated that larger FORs would result in more total time spent searching to locate task-specific AOIs. The Bayesian analysis reported a Bayes Factor of .021, providing insufficient evidence to support the hypothesis. Mauchly's Test of Sphericity was non-significant, W = .885, p = .125.

ANOVA Results

Repeated measures ANOVAs were conducted with corrections to address violations of sphericity and to explore mean differences across FOR conditions for each dependent variable. All of the ANOVA results mirrored those of the Bayesian analyses, with no significant differences found between FORs.

Hypothesis 1 (Task Performance)
Task performance was expected to vary with FOR size, predicting superior results in smaller FORs. When applying Greenhouse-Geisser correction for violation of sphericity, the results were non-significant with corrected degrees of freedom $F(1.572, 27.51) = .180$, $p = .283$, $\eta^2 = .005$. This suggests that changes in FOR size did not significantly affect task performance.

Hypothesis 2 (Seat Kit Release Accuracy)
It was predicted that larger FORs would yield more accurate seat kit release altitudes. When applying the Greenhouse-Geisser correction for violation of sphericity, the analysis showed no significant effects, with $F(1.951, 34.15) = .047$, $p = .954$, $\eta^2 = .001$. These results suggest that difference in FOR size did not significantly affect the accuracy of seat kit release altitudes.

Hypothesis 3 (Course Correction Time)
Larger FORs were hypothesized to increase the course correction times after identifying a wind cue. The application of the Greenhouse-Geisser correct for violation of sphericity yielded non-significant results, with corrected degrees of freedom $F(1.916, 33.58) = .839$, $p = .432$, $\eta^2 = .023$. These findings indicate that changes in FOR size did not significantly impact course correction times as initially hypothesized.

Hypothesis 4 (Confidence in Training)
We expected that larger FORs would improve confidence in performance. When applying the Greenhouse-Geisser correction for violation of sphericity, the results were not significant, with corrected degrees of freedom $F(1.342, 22.81) = 3.339$, $p = .062$, $\eta^2 = .087$. These findings suggest that FOR size did not significantly impact confidence in performance.

Hypothesis 5 (Time to First Fixation)
To determine if smaller FORs lead to longer times to first fixation, as predicted, an ANOVA was run with a Greenhouse-Geisser correction for violation of sphericity. This analysis yielded non-significant results, with corrected degrees of freedom $F(1.8, 30.60) = 1.182$, $p = .310$, $\eta^2 = 0.033$. These findings suggest that smaller FOR sizes do not significantly affect the times to first fixation as initially hypothesized.

Hypothesis 6 (Dwell Time)
Larger FORs were expected to yield longer dwell times. When applying the Greenhouse-Geisser correction for violation of sphericity, the results were non-significant with corrected degrees of freedom $F(1.881, 32) = .130$, $p = .866$, $\eta^2 = 0.004$. These findings indicate that larger FOR sizes do not significantly affect dwell times as initially hypothesized.

Hypothesis 7 (Number of Fixations)
Larger FORs were also expected to yield a smaller number of fixations. With the application of the Greenhouse-Geisser correction for violation of sphericity, the results were non-significant, showing corrected degrees of freedom $F(1.575, 26.775) = .850$, $p = .409$, $\eta^2 = .024$. These findings indicate the changes in FOR size do not significantly influence the number of fixations as initially hypothesized.

Hypothesis 8 (Time Searching)
This hypothesis stated that larger FORs would result in more total time spent searching to locate task-specific AOIs. The application of the Greenhouse-Geisser correction for violation of sphericity resulted in non-significant findings, with corrected degrees of freedom $F(1.794, 30.50) = .490$, $p = .595$, $\eta^2 = .014$. These results suggest that differences in FOR size do not significantly impact the total time spend searching for task-specific AOIs as initially hypothesized.

DISCUSSION

The study's utilization of both Bayesian analyses and repeated measures ANOVA provides a comprehensive examination of the effects of field of regard (FOR) variations on VR parachute descent simulation outcomes. Bayesian analysis primarily assessed the strength of evidence against the null hypothesis across various performance and perceptual metrics. The Bayesian results consistently indicated insufficient evidence to reject the null hypothesis for all tested outcomes, suggesting that varying FOR sizes did not markedly affect the performance measures evaluated. This is quantitatively supported by Bayes Factors significantly less than 1 across all hypotheses, highlighting a stronger likelihood of the observed results under the null hypothesis rather than the experimental hypothesis. Further, the ANOVA results largely mirrored those of the Bayesian analysis, with no significant mean differences found across most FOR conditions, further reinforcing the Bayesian findings.

Future research into parachute descent training for emergency situations could expand these findings in several areas. First, research into how individual differences (e.g., gaming interest, VR experience levels) may influence learning and retention, and may provide insights into training programs. Additionally, with technical advances such as foveated rendering, there may be future considerations for adaptive VR systems that dynamically adjust the FOR based on real-time performance data or offer alternative instructional scaffolding techniques. Similarly, personalized training protocols informed by individual learning styles and skill levels may offer a tailored approach to optimize training outcomes and maximize participant performance. Finally, conducting longitudinal studies could provide data for evaluating the lasting impact and transferability of skills acquired through PDT training, providing invaluable insights into the effectiveness of such training methodologies over time.

REFERENCES

Abich IV, J., Parker, J., Murphy, J. S., & Eudy, M. (2021). A review of the evidence for training effectiveness with virtual reality technology. *Virtual Reality, 25*(4), 919-933. https://doi.org/10.1007/s10055-020-00498-8

American Psychological Association (2018, April 19). *APA dictionary of psychology: Spatial orientation*. https://dictionary.apa.org/spatial-orientation

Angel-Urdinola, D. F., Castillo-Castro, C., & Hoyos, A. (2021). *Meta-analysis assessing the effects of virtual reality training on student learning and skills development*. World Bank Group. https://documents1.worldbank.org/curated/en/204701616091079027/pdf/Meta-Analysis-Assessing-the-Effects-of-Virtual-Reality-Training-on-Student-Learning-and-Skills-Development.pdf

Bowman, D. A., & McMahan, R. P. (2007). Virtual reality: How much immersion is enough? *Computer*, *40*(7), 36–43. https://doi.org/10.1109/mc.2007.257

Clifford, R. M., McKenzie, T., Lukosch, S., Lindeman, R. W., & Hoermann, S. (2020). The effects of multi-sensory aerial firefighting training in virtual reality on situational awareness, workload, and presence. *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 93-100. https://doi.org/10.1109/VRW50115.2020.00023

Guilford, J. P., & Zimmerman, W. S. (1948). The Guilford-Zimmerman aptitude survey. *Journal of Applied Psychology, 32*(1), 24–34. https://doi.org/10.1037/h0063610

Hartley, C. A., Hurley, K. J., Morganson, V. J., Kass, S. J., Wheeler Atkinson, B. F., & Pierce, M. B. (2022). Eject, Eject, Eject! Conducting a Cognitive Task Analysis to Assess Parachute Descent Training Simulators. *Ergonomics in Design*, 10648046221124786.

Hepperle, D., & Wolfel, M. (2023). Similarities and differences between immersive virtual reality, real world, and computer screens: A systematic scoping review in human behavior studies. *Multimodal Technologies and Interaction, 7*(6), 56. https://doi.org/10.3390/mti7060056

Howard, M. C., Gutworth, M. B., & Jacobs, R. R. (2021). A meta-analysis of virtual reality training programs. *Computers in Human Behavior*, 121. https://doi.org/10.1016/j.chb.2021.106808

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.

Kyritsis, M. & Gulliver, S. R. (2009). Gilford Zimmerman orientation survey: A validation. Proceedings of the 7th International Conference on Information, Communications and Signal Processing (ICICS). https://doi.org/10.1109/ICICS.2009.539759

McMahan, R. P. (2011). *Exploring the effects of higher-fidelity display and interaction for virtual reality games* [Doctoral dissertation, Virginia Tech]. Virginia Tech Electronic Theses and Dissertations. http://hdl.handle.net/10919/30123

Moroney, W. F., & Lilienthal, M. G. (2009). Human factors in simulation and training: An overview. In D. A. Vincenzi, J. A. Wise, M. Mouloua, & P. A. Hancock (Eds.), *Human factors in simulation and training* (pp. 3-33). Taylor & Francis.

Narciso, D., Melo, M., Rodrigues, S., Paulo Cunha, J., Vasconcelos-Raposo, J., & Bessa, M. (2021). A systematic review on the use of immersive virtual reality to train professionals. *Multimedia Tools and Applications, 80*, 13195-13214. https://doi.org/10.1007/s11042-020-10454-y

Peracchio, A. S. (2020). *Training with virtual reality: The role of self-efficacy and perceived spatial ability* [Master's thesis, Xavier University]. OhioLINK Electronic Theses and Dissertations Center. http://rave.ohiolink.edu/etdc/view?acc_num=xupsy1593428888934927

Radhakrishnan, U., Koumaditis, K., & Chinello, F. (2021). A systematic review of immersive virtual reality for industrial skills training. *Behaviour & Information Technology, 40*(12), 1310-1339. https://doi.org/10.1080/0144929X.2021.1954693

Ragan, E. D., Kopper, R., Schuchardt, P., & Bowman, D. A. (2013). Studying the effects of stereo, head tracking, and field of regard on a small-scale spatial judgment task. *IEEE Transactions on Visualization and Computer Graphics, 19*(5), 886-896. https://doi.org/10.1109/TVCG.2012.163

Stanney, K. M., Nye, H., Haddad, S., Hale, K. S., Padron, C. K., & Cohn, J. V. (2021). Extended reality (XR) environments. In G. Salvendy, & W. Karwowski (Eds.), *Handbook of human factors and ergonomics* (5th ed., pp. 782-815). Wiley.

Strojny, P., & Dużmańska-Misiarczyk, N. (2023). Measuring the effectiveness of virtual training: A systematic review. *Computers & Education: X Reality*, 2. https://doi.org/10.1016/j.cexr.2022.100006

Vive. (2018, March 19). *HTC VIVE announces price of Vive Pro HMD at $799, pre-orders start today; price of Vive reduced to $499*. https://www.vive.com/us/newsroom/2018-03-19-2/

Vive. (n.d.). *Vive Pro*. https://web.archive.org/web/20180503190026mp_/https://www.vive.com/us/product/vive-pro/

Xie, B., Liu, H., Alghofaili, R., Zhang, Y., Jiang, Y., Lobo, F. D., Li, C., Li, W., Huang, H., Akdere, M., Mousas, C., & Yu, L. F. (2021). A review on virtual reality skill training applications. *Frontiers in Virtual Reality*, 2. https://doi.org/10.3389/frvir.2021.645153