Modern Training Practices: Studying the Use of Virtual Reality for Training Procedures in Flightcrew Training

Final Technical Report and Results of Phases 2 & 3 Studies

(With revisions through 27 December 2023)

Lead authors:
Blake Nguyen, Nathan Sonnenfeld, Fiona Duruaku, and Florian Jentsch

With assistance from:
Andy Silva, Caroline Gomez, and Alex Alonso

Principal Investigator (Institution):	Florian Jentsch (University of Central Florida, Orlando)
Research Title:	Modern Training Practices
Grant/CRA Number:	FAA CRA 692M151940002
Period of Performance:	9/13/19 – 9/30/23
Research Requirement:	Pilot Training and Checking Methods (A11G.HF.1)
Technical Sponsors:	Kathy Abbott (AVS), Joshua Jackson, Lori Thomasson (AFS-280)
Program Manager:	Chuck Perala (ANG-C1)
Report Date:	30 September 2023 (Revised 27 December 2023)

TECHNICAL REPORT DOCUMENTATION PAGE

1. REPORT NO.	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
	N/A	N/A

4. TITLE AND SUBTITLE
Modern Training Practices: Studying the Use of Virtual Reality for Training Procedures in Flightcrew Training

5. REPORT DATE
Dec 2023

6. PERFORMING ORGANIZATION CODE
N/A

7. AUTHOR(S)
Blake Nguyen, https://orcid.org/0000-0002-2826-2171
Nathan Sonnenfeld, https://orcid.org/0000-0001-5233-1272
Fiona Duruaku, https://orcid.org/0009-0008-8035-2183
Florian Jentsch, https://orcid.org/0000-0001-9879-7764

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. PERFORMING ORGANIZATION NAME AND ADDRESS
The University of Central Florida Board of Trustees
12201 Research Parkway, Suite 501
Orlando, FL 32826-3246

10. WORK UNIT NO.
N/A

11. CONTRACT OR GRANT NO.
Cooperative Agreement 692M151940002 Amendment 003

12. SPONSORING AGENCY NAME AND ADDRESS
Federal Aviation Administration
NextGen Human Factors Division
800 Independence Avenue, SW
Washington DC, 20591

13. TYPE OF REPORT AND PERIOD COVERED
Final Report
Oct 2019- Dec 2023

14. SPONSORING AGENCY CODE
ANG-C1

15. SUPPLEMENTARY NOTES
This report has been modified by the authors from the contractual final deliverable, for improved clarity for a general audience. The original report was funded through Federal Aviation Administration Aviation Research Grants Program, Cooperative Agreement 692M151940002. "Modern Training Practices: Studying the Use of Virtual Reality for Training Procedures in Flightcrew Training", Deliverable: Final Technical Report and Results of Phases 2 & 3 Studies. The prime vendor for the research was The University of Central Florida. The principal investigator was Dr. Florian Jentsch.

16. ABSTRACT
Flightcrew training is a cornerstone of aviation safety, ensuring that pilots and other crewmembers possess the skills and knowledge required to navigate complex aircraft and respond effectively to non-normal situations. As technology continues to advance and new training demands continue to emerge in the aviation industry, the use of modern training practices will become more predominant. It is therefore important to have an understanding of how modern training practices—in particular, virtual reality (VR), mixed reality (MR), augmented reality (AR), or other immersive technologies, collectively referred to as extended reality (XR)—are best implemented in air carrier training. Thus, our research was aimed at (a) gathering and reviewing information related to the current state of the art of modern training practices and methodologies, (b) identifying how air carriers are currently implementing and evaluating these training approaches, and (c) providing information on potential future uses. Specifically, the goal was directed towards identifying and testing emergent technologies that could be suitable in specific areas of flightcrew training. The results of this research could serve to inform Federal Aviation Administration (FAA) personnel who update regulatory and guidance material which relate to air carrier pilot training, procedures, and operations. Data may also inform Flight Standards Service (FS) personnel who evaluate, approve, and oversee flight deck operations, procedures, and operator training programs.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
human factors; pilot training; modern training practices; pilot training and checking methods;	No restrictions.

virtual reality for training procedures; flightcrew training; NextGen; aviation flight deck research			
19. SECURITY CLASSIF. (OF THIS REPORT) Unclassified	20. SECURITY CLASSIF. (OF THIS PAGE) Unclassified	21. NO. OF PAGES 78	22. PRICE N/A

Form DOT F 1700.7 (08/72)

Acknowledgements

This research was completed with funding from the Federal Aviation Administration (FAA) NextGen Human Factors Division (ANG-C1) in support of the FAA Aviation Safety (AVS) organization. We thank our FAA program manager, Dr. Chuck Perala (ANG-C1), and technical sponsors, Dr. Kathy Abbott (AVS), Joshua Jackson, and Lori Thomasson (both AFS-280), as well as FAA Contract Support Personnel (Eddie Austrian and Kevin Siragusa). We would also like to thank the subject matter experts (SMEs) from government, industry, and academia who, through conversations held as a part of scientific & technical meetings, through committee work, or in other circumstances, informed the development of this document.

Acronyms

Acronyms Related to Human Factors & Aviation Terminology

2D	Two Dimensional
3D	Three Dimensional
AC	Advisory Circular
AI	Artificial Intelligence
ANOVA	Analysis of Variance
APD	Aircrew Program Designee
AR	Augmented Reality
CAVE	Cave Automatic Virtual Environment
CBT	Computer-Based Training
CFR	Code of Federal Regulations
CRM	Crew Resource Management
E&DL	Electronic & Distance Learning
FFS	Full Flight Simulator
FO	First Officer
FSTD	Flight Simulation Training Device
HMD	Head-Mounted Display
I-VR	Immersive Virtual Reality (see also Virtual Reality)
IPD	Inter-Pupillary Distance
LCP	Line Check Pilot
MR	Mixed Reality
PC	Personal Computer
SBT	Scenario-Based Training
SME	Subject Matter Expert
SSQ	Simulation Sickness Questionnaire
TEE	Training Effectiveness Evaluation
VR	Virtual Reality
XR	Extended Reality

Acronyms Related to Study-Specific Terminology

FlightPET	Flight Procedures Experimental Training testbed
T1	Time Point 1
T2	Time Point 2

Acronyms Related to Organizations

ACT ARC	Air Carrier Training Aviation Rulemaking Committee
ARL	Army Research Laboratory
FAA	Federal Aviation Administration
FS	FAA Flight Standards Service
EKT WG	ACT ARC Electronic Knowledge Training Working Group
HFES	Human Factors & Ergonomics Society

IST	UCF Institute for Simulation & Training
UCF	University of Central Florida

Acronyms Related to Conferences & Meetings

AAM	Advanced Air Mobility
AFWERX	Air Forcc WERX
AVID	UCF AR/VR Innovation Discovery symposium
BCI	Brain-Computer Interface
CAT	Civil Aviation Training
GIFT	Generalized Intelligent Framework for Tutoring
I/ITSEC	Interservice/Industry Training, Simulation and Education Conference
ISAP	International Symposium on Aviation Psychology
NEAT	FAA New and Emerging Aviation Technologies (NEAT) series
NSF	National Science Foundation
NTSA	National Training & Simulation Association
SIW	Simulation Innovation Workshop
WATS	World Aviation Training Summit

Modern Training Practices: Studying the Use of Virtual Reality for Training Procedures in Flightcrew Training

Executive Summary

Why We Did This Study

Flightcrew training is a cornerstone of aviation safety, ensuring that pilots and other crewmembers possess the skills and knowledge required to navigate complex aircraft and respond effectively to non-normal situations. As technology continues to advance and new training demands continue to emerge in the aviation industry, the use of modern training practices will become more predominant. It is therefore important to have an understanding of how modern training practices—in particular, virtual reality (VR), mixed reality (MR), augmented reality (AR), or other immersive technologies, collectively referred to as extended reality (XR)—are best implemented in air carrier training. Thus, our research was aimed at (a) gathering and reviewing information related to the current state of the art of modern training practices and methodologies, (b) identifying how air carriers are currently implementing and evaluating these training approaches, and (c) providing information on potential future uses. Specifically, the goal was directed towards identifying and testing emergent technologies that could be suitable in specific areas of flightcrew training. The results of this research could serve to inform Federal Aviation Administration (FAA) personnel who update regulatory and guidance material which relate to air carrier pilot training, procedures, and operations. Data may also inform Flight Standards Service (FS) personnel who evaluate, approve, and oversee flight deck operations, procedures, and operator training programs.

What We Found

Phase 1 Literature Review

The goal for this research phase was to evaluate the use and effectiveness of electronic and distance learning (E&DL), broadly defined, for training procedures in the aviation domain. Our first step was a review of the extant literature. We reviewed over 1,200 academic, technical, and selected industry documents across relevant domains, including aviation, other transportation sectors, defense, and healthcare. This scientific and technical information and data helped us identify the effectiveness, advantages, and limitations of using modern E&DL technologies and approaches for training procedures, and allowed us to assemble other information related to the design, development, and evaluation of these modern training configurations (Sonnenfeld et al., 2021b). We also developed a compendium glossary and database of over 3,000 terms and definitions related to modern training that we made available to the FAA Air Carrier Training - Aviation Rulemaking Committee (ACT-ARC) Electronic Knowledge Training Working Group (EKT WG), and which became referenced as part of ACT ARC *Recommendation 20-11: Terms and Definitions related to eLearning.*

Our key findings and considerations, based on our Phase 1 review (Sonnenfeld et al., 2021b), were:

- Modern E&DL configurations, when designed well, have the potential to be equally as, or even more effective than, exclusively traditional (e.g., lectures, presentations, online learning modules) training methods, specifically for objectives supporting the training of flightcrew-relevant procedures.
- The literature generally supported equivalence of new training in procedures training across different delivery technologies and training approaches. That is, how training content is designed may have more of an impact on outcomes than the delivery medium.
- Certain configurations of technologies and approaches may provide substantial gains in specific training outcomes, although definitive guidelines for pairing modern configurations to training outcomes were not found.
- Procedures training should encompass both declarative knowledge (i.e., “know this”) and procedural skill (i.e., “know how”) training and practice.
- Effective E&DL configurations for declarative and procedural knowledge training often include instructional features such as multimedia, interactivity, and intelligent tutoring. Effective E&DL configurations for skill training generally use some form of interactive demonstration and/or simulation.
- To expand the benefits of simulation, scenario-based training (SBT) should be designed to incorporate elements of intelligent systems.
- There appears to be promising but limited evidence that serious games—simulations developed to support learning/training objectives while leveraging game mechanics and elements—may be effective for procedures training.
- E&DL using mobile devices is best suited for short-duration training content (microlearning). Mobile devices may also be best used to provide training at the point and/or time of need (training-on-demand).
- Well-designed extended reality (XR, encompassing Virtual Reality – VR, Mixed Reality – MR, and Augmented Reality – AR) can be as effective for improving procedures training outcomes as conventional training for the knowledge portions and may be even more suitable than conventional training for physical tasks. Use of extended reality configurations should be determined by other than effectiveness alone.
- There is substantial evidence to suggest that VR technologies can be effectively used to deliver procedures training in aviation.
- There is less evidence that AR training can be effective for training procedures, and safety concerns for use of AR for flightcrew training in the aircraft remain to be addressed.
- VR shows promise as a technology that may revolutionize training—yet developers have struggled to enable the systems to move from experimental or prototype settings and to integrate them into real-world training contexts.

Overall, the results of our literature review suggested that use of XR-solutions, and VR technologies in particular, may provide unique benefits to support the training of flightcrew-relevant procedures. The most important challenge to implementing XR for training procedures is selecting an appropriate configuration among available technologies and approaches. A useful three-dimensional framework to conceptualize, classify, and compare XR solutions considers *immersion* and *interactivity*, as originally proposed by Slater and Wilbur (1997), and adds *realism* as a third dimension. In this context, immersion refers to the degree to which the XR solution provides a sense of vivid presence in the target environment; interactivity to the way the experience is changed from passive to active; and realism describes how the XR experience approximates the existing reality of the target environment.

Phase 2 & 3 Research Objectives and Methods

Having recognized the potential of XR simulations, our Phase 2 and Phase 3 research sought to investigate *immersion* and *interactivity* using various simulation environments and instructional elements. Our aim was to understand the degree to which interactivity might compensate for immersion, especially for tasks that have different levels of involved motor actions.

To obtain answers to these questions, in Phases 2 and 3 of this project, we compared different instructional elements in two empirical studies, with similar goals: In both studies, we compared more immersive implementations of training simulations (VR using a head-mounted display) versus less immersive ones (personal computer with monitor, or tablet). We collected perceptions of usability, assessments of instrumentality and efficacy, measures of learning, and indications of cybersickness from participants who interacted with either immersive VR in pilot-relevant procedural aviation tasks, or their PC- or tablet-based equivalents. We also engaged in detailed comment capturing. For example, we asked the study participants to provide their thoughts on the training that they received, what they liked or disliked about the training, and around what time or point did they feel familiar or comfortable with the VR system, in addition to taking standardized quantitative measures.

In Phase 2, a total of 102 participants (college students) were randomly assigned to one of four conditions in a 2 (Immersion; lower vs. higher) x 2 (Interactivity; lower vs. higher) between-subject factorial design. The Phase 2 study focused on the impact of different instructional strategies such as more passive (i.e., static text and media) vs. more active (i.e., dynamic interactions, object enrichment, and cues and clues) interactions, and crossed them with immersion (i.e., immersive VR vs PC-monitor). The preflight exterior inspection task simulation and training system used in this study was developed in-house at the University of Central Florida (UCF) Institute for Simulation & Training (IST) as part of this research project and featured a large number of novel instructional approaches across both head-mounted display (HMD) VR and personal computer (PC) monitor displays. It was intended to explore how individuals acquire the necessary knowledge and skills to perform a lengthy procedure with many sub-steps during air carrier-level operations, yet be accessible to a broader study population.

In Phase 3, we collected data from 30 active Part 121 air carrier pilots at a major US airline training center. The pilots were randomly assigned to one of four study conditions, in a 2 (immersion, high vs. low) [between] x 2 (procedure order; exterior preflight inspection first vs. initial preflight setup/flows first) [between] x 2 (simulations: exterior preflight inspection vs. initial preflight setup/flows) [within] mixed-model design. The air carrier pilots each interacted with two different training implementations, (i) the same trainer for external preflight inspections as in the Phase 2 study and (ii) an externally developed trainer for initial preflight setup/flows. We compared immersive (VR with HMD) and less immersive (PC with monitor/tablet) implementations. We aimed to understand pilots’ perceptions of training using immersive VR vs desktop/tablet, to compare the applications and implementations, and to obtain quantitative data on instrumentality, perceived effectiveness, usability, as well as simulator sickness.

Phase 2 & 3 Research Results and Discussion

The results of our analyses of both quantitative data and qualitative comments from both studies showed that both the highly immersive VR- and the less immersive PC-/tablet-based equivalents of the training simulations were generally assessed very positively by both participant groups. Both training simulations were considered improvements over currently available training methods (e.g., a static flight deck mock-up [“paper tiger”] or using handbooks or training texts for the initial preflight setup/flows; CBT with videos or static PowerPoints for the exterior preflight inspection) by the Phase 3 pilot participants, but there were notable differences in usability, perceived instrumentality, and the level of discomfort (cybersickness) experienced by our study participants between implementations, depending on the task.

The *preflight exterior inspection* simulation, regardless of implementation, was associated with substantial knowledge gains in the Phase 2 study, when compared to the pre-study level of knowledge among participants. Further, there were no significant differences in either gains or absolute post-test knowledge measures between the VR and the PC implementations. However, we observed differences in usability. The VR solution was rated lower, on average, albeit this difference was not statistically significant. We also observed differences in cybersickness, such that the VR solution led to, on average, highly significant and substantially greater levels of cybersickness (doubling SSQ scores from below 20 to near 40, at a medium statistical effect size).

A similar pattern was found among the air carrier pilots, who were also very positive towards both implementations of the preflight exterior inspection software. However, the air carrier pilots also indicated that they thought the VR implementation was more cumbersome, probably not needed, and they reported nearly four-times the levels of cybersickness (albeit at very low absolute levels) after experiencing the VR-implementation for only 10-15 minutes, when compared with the PC-implementation. The Phase 3 pilot participants also rated the desktop implementation of the preflight exterior inspection software higher than other currently used alternatives, such as videos, PowerPoint presentations, etc. Together, this indicates that, as we predicted, for a procedural task that had a perceptual but limited motor component, the use of immersive VR was not justified in terms of acceptance, perceived instrumentality, or effectiveness. Indeed, using immersive VR for this task likely would be contraindicated, due to higher levels of frustration and cybersickness that trainees may experience, compared to the PC-desktop equivalent.

In contrast, for the *initial preflight setup/flows* training software, which targeted a task with a very active and spatially sensitive motor component, participants in the Phase 3 air carrier study consistently expressed their belief that the immersive VR-implementation was “value added”, not only over the tablet-based version, but even when compared to a full-flight simulator (FFS) or any other form of training device, from FTDs to the cardboard flight desk poster static mock-up (i.e., “paper-tiger”). Despite the fact that those using the VR implementation, on average, also rated their cybersickness levels about 4 times the level of those who interacted with the tablet version (albeit at very low absolute levels) and despite their concerns about having to learn the VR technology, the ability to actively interact with the flight deck, reach for and operate switches, knobs, and buttons on the overhead panel, glare shield, etc., was consistently considered a major improvement for this highly spatial and active task. Pilots referred to it as "game changing," with one summarizing the sentiment as "Every pilot wants to have this in [their] hotel room, yesterday." Still, the tablet version was also very highly rated, based on its interactivity and the opportunity to practice the initial preflight setup/flows anywhere and anytime. In fact, pilots rated the CBT (tablet) version of the training, after experiencing it, as being as useful for practice as the Full Flight Simulator, especially for initial training and for quick review.

Project Conclusions and Practitioner Findings

The results of this research provided additional evidence that virtual simulations might be suitably used to augment or potentially replace existing approaches for training the targeted tasks in an air carrier setting. Specifically, immersive virtual simulations implemented in XR, especially for highly spatial tasks, might be used to provide lower-fidelity simulation of "busy work" tasks, in lieu of more expensive FTDs, FSTDs, or even FFSs. Most participants mentioned that VR would be most appropriate for new hire and initial transition/differences training. Furthermore, spatial interactivity was identified as a key advantage of the immersive VR over existing training, such as using a paper tiger, desktop, and tablet training. Finally, for VR to be acceptable to flightcrews, findings suggested that it should have task fidelity, high visual quality, provide multimodal feedback, and include effective tutorials to enhance familiarization.

Based on our empirical observations across all three phases, we developed a set of 35 actionable findings related to the design, introduction, and use of immersive technologies in the context of flightcrew training. These findings, and the observations they were based on, are further elaborated on in the Discussion, in sections 5.2. through 5.5.

In the following, we present some key findings from among the 35. They are grouped below within topic areas; the "Finding [#]" at the end of each finding refers to the list of findings in the Discussion section (5.2 through 5.5).

Potential Benefits of XR/VR for Aviation Training

XR-environments have the potential to provide equal or better training for applied aviation tasks than current training conducted in other environments, at potentially lower cost, higher scalability, and greater access and customization to trainees. [*Finding 1*]

Highly immersive XR solutions allow relevant training environments to be experienced anywhere-anytime, while alleviating the dependence on large, fix-based infrastructure. [*Finding 8*]

Highly immersive XR applications present new opportunities for joint training of physically distributed trainees and trainers, who access the training environment and interacting with it synchronously, but from very different physical locations. [*Finding 9*]

Highly interactive XR simulations that include novel instructional elements, such as dynamic visualizations, multi-media, etc. may be preferable over training in higher-fidelity devices which do not include those instructional elements. This also can free up high-fidelity devices, such as Full Flight Simulators, for other training and practice opportunities. [*Finding 24*]

Limitations & Trade-Offs of XR/VR in Aviation Training

An XR-solution that is poorly designed, employed, or assessed, may not reach its potential; indeed, it may be worse than the traditional training it is intended to supplement or replace. [*Finding 2*]

XR and XR-environments are not training, *per se*; they are a medium for conveying information to pilots. To reach their potential for training, XR-solutions need to be carefully integrated into a systematically developed, used, and assessed training eco-system. [*Finding 3*]

Highly immersive XR solutions may make interactions in the real world between trainers and/or trainees more difficult while trainees are immersed in VR. [*Finding 11*]

The lack of accurate avatars with realistic representations of appearance, nonverbals, and facial/gaze cues currently limit the utility of VR for tasks involving crew coordination and communication. [*Finding 12*]

The need to recalibrate XR equipment periodically can reduce presence, increase frustration, and induce or exacerbate cybersickness. [*Finding 13*]

Using XR solutions today still requires a lot of specialized know-how that many trainers and trainees do not have yet. This leads to "VR-overhead" needed for simply using an XR training solution and can reduce the technology's effectiveness. [*Finding 25*]

XR is still in an early stage of technology development, characterized by a lack of standardization and consolidation. As a result, each combination of hard- and software is somewhat bespoke. This hinders assessing and comparing solutions. It also makes it difficult to generalize results and provide minimum standards for XR solutions. [*Finding 26*]

Mechanisms & Design Considerations for XR/VR

A useful way to conceptualize, classify, and compare XR solutions is to assess and quantify them along three dimensions: immersion, interactivity, and realism. [*Finding 4*]

There are two aspects of interactivity relevant for XR systems: the degree to which interactions with the training environment are possible and consequences are displayed (the "what"), and the way in which these interactions are implemented (the "how"). The "what" is at least as important as the "how". [*Finding 14*]

High interactivity is not only highly desired by users, but it also improves learning, especially under conditions of low immersion. [*Finding 15*]

When trade-offs must be made between immersion or interactivity, increasing interactivity may be able to "make up" for loss in immersion. [*Finding 16*]

How interactions with virtual objects are implemented in VR is currently not standardized. Small differences in implementation may be associated with large differences in usability. [*Finding 20*]

Implementing XR/VR Training

Instructional support is critical for the successful use of an XR simulation in training. New users of XR technologies desire, and benefit from, being guided and aided by a human trainer when using XR training systems, especially at initial use. [*Findings 29 and 30*]

Training organizations interested in implementing XR training should carefully consider how and where to introduce the training. A phased approach, with limited small-scale trials, potentially on a volunteer basis, that collects data on usability, availability, and acceptance of the technology, can provide important information to increase acceptance and success. [*Finding 31*]

Cybersickness in XR/VR

Most people are to some degree susceptible to cybersickness, and, as both the extant literature and our studies demonstrate, about 10% to 15% of the population are particularly susceptible. Training organizations should be cognizant that highly susceptible individuals may be found in any group of XR-users. [*Finding 32*]

Training organizations should have steps in place to minimize, assess, and respond to cybersickness. [*Finding 33*]

There are general principles that can reduce the likelihood, incidence, and magnitude of cybersickness in VR (see table in Section 5.5). However, it is the impact of their *unique combination* that is difficult to model or predict; thus, user testing is critical. [*Finding 34*]

Table of Contents

List of Figures

List of Tables

Modern Training Practices: Studying the Use of Virtual Reality for Training Procedures in Flightcrew Training

1.0. Background

Over the past two decades, asynchronous online training interventions delivered using computers and even mobile devices (e.g., tablets) have become commonplace as a supplement or, in some cases, as a replacement for flightcrews' traditional classroom and simulator training (e.g., Saraçyakupoğlu, 2020). Various other modern learning technologies have also been highlighted for their potential applicability within air carrier training configurations (Sonnenfeld et al., 2021a), including intelligent systems, conversational agents, and serious games (Aleven et al., 2013; Azevedo & Gašević, 2019; Graesser, 2013). However, technologies along the spectrum of extended reality (XR; Rauschnabel et al., 2022) in general—and VR configurations in particular—have been of particular interest, because of their potential role in fulfilling specific niches within the flightcrew training cycle. For example, VR has been proposed as a supplement to classroom training, as a bridge to rehearse procedural skills and tasks before simulator training, or as a means to provide enrichment or extracurricular practice following simulator training (Schaffernak et al., 2022; see also Brusso et al., 2014).

There are a number of advantages, but also some limitations, related to the use of VR configurations for training in aviation, commonplace in discussions at industry training conferences. Among these are:

- VR configurations show promise as a technology that may revolutionize training—yet developers have struggled to enable the systems to move from experimental or prototype settings and to integrate them into real-world training contexts.
- VR configurations may fit an unmet niche to bridge training ranging from online learning and paper instrument panels to simulator training and live practice—yet it is not yet shown that they show the fidelity, reliability, and effectiveness necessary to warrant any redistribution of training investment.
- VR configurations seem the ideal tool to engage the next generation of pilots—yet popular conceptions of generational differences have given decision-makers pause to consider whether VR technologies may alienate older pilots and thus may not be appropriate for trainees across all demographics.
- Finally, while marketing would suggest VR configurations are ideal to support training transfer, these qualities remain poorly understood, and there is a valid concern of whether such qualities outweigh their current limitations (e.g., cybersickness, technical issues, cost).

In Phase 1 of this research, we found that there has been some, including empirical, support in the literature for the use of VR as a delivery medium for flightcrew training, particularly for the training of flight crew procedures (e.g., Sonnenfeld, Nguyen, Boesser, & Jentsch, 2021). The literature suggests that training for procedures using VR can be *at least* equivalent in effectiveness to training delivered via other mediums, although the "value added" of VR is not yet clear for training procedures, when compared with other forms of training using computerized simulations (Sonnenfeld et al., 2021a; 2021b). Numerous studies in the literature have shown that trainees' individual differences (e.g., experience, prior knowledge, gender, etc.), systems' instructional characteristics (e.g., interactivity, instructor involvement, instructional strategies and techniques), task characteristics (e.g., the domain and complexity of targeted competencies), and specific system characteristics (e.g., immersion, fidelity), impact training effectiveness, including those leveraging VR configurations.

One deficiency in the empirical literature and in extant industry case studies has been an insufficient analysis of the interaction of these factors (e.g., instructional characteristics) and the system characteristics underlying the training medium (e.g., VR) on training effectiveness, in the context of specific topics and competencies relevant to flightcrews. Furthermore, there has been a deficiency of research on the types of instructional elements which may be effective for specific competencies and across different types of instructional events. This has resulted in conflicting recommendations for the design of training and learning interventions (Koedinger et al., 2012).

The results of this research could serve to inform Federal Aviation Administration (FAA) personnel who update regulatory and guidance material which relate to air carrier pilot training, procedures, and operations. Data may also inform Flight Standards Service (FS) personnel who evaluate, approve, and oversee flight deck operations, procedures, and operator training programs. This information and data may be used by these personnel to inform the safe integration of technology and procedures in flight operations and training and inform potential updates to regulations and guidance reflecting contemporary training practices within the civil aviation training industry.

1.1. Purpose

We conducted a series of research activities to investigate the suitability of modern training practices, particularly VR, for topics and tasks relevant to air carrier training. Our aim was to review and gather information related to modern training practices and methodologies, to identify how air carriers implement and evaluate the effectiveness of these training approaches, and how they are evaluating the effectiveness of their training programs. We sought to identify and test emerging technologies (e.g., VR) and distance learning techniques (e.g., computer-based training, mobile learning) that could be suitable in specific areas of flightcrew training. Our objectives were to:

a. Gather information from stakeholders in the air carrier training industry regarding their application of modern training practices and assessment methods (Phase 1);
b. Review the extant literature to identify modern training trends with relevance for flightcrew training (Phase 1);
c. Conduct baseline experimental studies to determine if there may be significant differences in training potential for task-based training using VR versus a PC-based simulation (Phase 2);
d. Conduct experimental studies using air carrier pilots at a collaborating carrier to examine the potential of VR for task-based training and comparing VR versus non-VR configurations in task-based training practices (Phase 3)
e. Provide information on successful ways to implement modern training practices (Final Technical Report); and

f. Provide recommendations to help decide which topics, skills, and knowledge in aviation training are appropriate for modern training practices or other learning methods (Final Technical Report).

The current report summarizes the results of our activities—in particular, we detail the findings of our investigation of how task characteristics (e.g., specific flightcrew procedures), system characteristics (e.g., immersion), instructional characteristics (e.g., interactivity), and individual characteristics (e.g., prior experience, age category, gender) impact indicators of training effectiveness (e.g., reactions and perceptions of learning).

Based on the analysis of data collected in two studies (Phase 2 in-house study and Phase 3 air carrier study), we add to the understanding of how immersion, interactivity and their combined influences affect the potential impact of VR technologies in modern flightcrew training. In particular, we obtained empirical data, ranging from trainees' reactions to perceptions of instrumentality and usability, to incidence of cybersickness, to selected performance data, for two exemplar flightcrew procedures (i.e., an exterior preflight inspection of a transport category aircraft and its initial preflight setup/flow).

In our empirical studies, we were guided by the following overarching research question:

- How does the immersiveness of the system, the interactivity of instruction, and the interaction between these constructs influence the potential use and effectiveness of training for flightcrew procedures?

In this document, we start by summarizing our findings from our Phase 1 review and discussions, providing relevant background information on the concepts of *immersion* and *interactivity*. We follow this by providing a summary of our methodologies for the empirical studies conducted under Phase 2 (conducted in-house using a college student sample) and Phase 3 (conducted at an air carrier using a sample of the carrier's pilots). After each methodology, the results and discussion are presented. Finally, a general discussion and practical implications are provided, with the intent that the results of the research we conducted may serve to provide the technical information, data, and recommendations needed to inform (1) Federal Aviation Administration (FAA) personnel who update regulatory and guidance material which relate to air carrier pilot training, procedures, and operations, and (2) Flight Standards Service (FS) personnel who evaluate, approve, and oversee flight deck operations, procedures, and operator training programs, of contemporary training practices as these personnel work to ensure the safe integration of technology and procedures in flight operations and training.

2.0. Overview of Research Activities Conducted Under Phase 1

The goals of this research phase were (a) to gather and review technical information related to air carriers modern training practices and methodologies, (b) to examine the suitability of emerging technologies and distance learning approaches for specific areas of flightcrew training, and (c) to identify how these air carriers implement and evaluate the effectiveness of these practices. The specific objectives of Phase 1 were to conduct a literature review to identify modern training trends involving emerging technologies [e.g., virtual reality (VR), augmented reality (AR), gaming, mobile learning] that could support aviation training programs, and to identify examples, similarities, and differences among these modern training approaches. We were also tasked with collecting and comparing information from air carriers regarding their current training practices and assessment methods, to identify how modern training methods may be delivered locally or at a distance, to identify how air carriers may evaluate the effectiveness of modern training practices, and to identify which topics, skills, and knowledge in aviation training may be delivered using modern training approaches.

2.1. Phase 1: Literature Review and Participation in the ACT-ARC Effectiveness of Knowledge Training Working Group (EKT WG)

We participated in and supported activities of the Effectiveness of Knowledge Training (EKT) Working Group (WG) within the FAA/Industry Air Carrier Training Advisory and Rulemaking Committee (ACT-ARC) as part of Phase 1 of the project. We provided subject matter expertise, background information, and material input to WG discussions, documents, and recommendations. We participated in forty-one (41) conference calls and three (3) multi-day face-to-face meetings with the EKT WG. We provided the EKT WG with a compendium glossary and database of over 3,000 terms and definitions related to modern training, used by the EKT WG as the basis for *ACT ARC Recommendation 20-11: Terms and Definitions related to eLearning under ACT ARC Initiative #44*. We also provided the EKT WG with technical information on training needs analysis, training design, and training evaluation, including an overview presentation of different categories of training evaluation with a categorization of over twenty (20+) training effectiveness evaluation frameworks, used by the EKT WG as a basis for *ACT ARC Recommendation 21-11: eLearning Guidance and Systematic Design Methodologies under ACT ARC Initiative #44*. These activities allowed us to exchange information with air carrier training experts from industry and government regarding current and near-future trends in air carriers' expectations for the use, implementation, and evaluation of emergent technologies and practices for flightcrew training, while supporting the efforts of the EKT WG with scientific and technical information. Our contributions to the committee's work helped the EKT WG identify a need for standardization and dissemination of terminology, definitions, and suggested practices as appropriate within FAA regulatory and guidance materials.

2.2. Phase 1: Industry Discussions

With assistance from the FAA and from EKT WG members, we contacted, and then conducted informal discussions with, representatives from air carriers in positions associated with flightcrew training. The purpose of these informal discussions was to identify how electronic and distance learning (E&DL) technologies and approaches were being used, and to obtain insights into current and near-future training trends. We also met with a small number of current air carrier line pilots (approx. 8) to gain insights on their experiences with E&DL approaches and insights into contemporary and near-future training practices. Supporting these discussions, we participated in eight (8) relevant industry workshops and conferences between December 2019 and March 2021, including the 2021 Annual FAA Human Factors Review, the Global Airline Training & Simulation (G-ATS) Conference, the International Symposium on Aviation Psychology (ISAP), the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC), the World Aviation Training Summit (WATS), and others including the Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium, the Simulation Innovation Workshop (SIW), and five (5) webinars including presentations by the Civil Aviation Training (CAT) Leader Forum, Microsoft's HoloLens team, and the Air Force Research Laboratory's AFWERX. These activities allowed us to obtain aviation training practitioner insights into current and near-future trends in their expectations for the use, implementation, and evaluation of emergent technologies and practices for flightcrew training, supported with an understanding of training trends within aviation and other relevant industries.

Throughout the remainder of our period of performance (March 2021 to September 2023), we supplemented these industry discussions with attendance and/or participation in (5) five industry workshops and conferences including the FAA/Industry Human Factors Roundtable (2023), WATS (2022/2023), the Central Florida Immersive Technology Summit, and the National Science Foundation (NSF) Advanced Air Mobility (AAM) Conference, also attending webinars on modern training approaches and technologies (e.g., virtual reality, experiential learning, adaptive training, artificial intelligence) including series by the National Training & Simulation Association's (NTSA) Tech Grove, Halldale Group, and a webinar by Varjo/OpenBCI on the use of biometric data in VR/AR devices. As discussed later in this report, as a supplemental activity for our Phase 3 air carrier study, we concluded our industry discussions with a series of informal demonstrations and discussions with over thirty (30) subject-matter experts (SMEs) representing a diverse range of stakeholders (e.g., management, instructional design, multimedia) from an air carrier training department pertaining to the prospective application of modern training approaches (e.g., virtual reality, mobile learning) in flightcrew training. These discussions helped to inform our Phase 2 and Phase 3 research activities, as well as the discussion and conclusions of this report.

2.3. Phase 1: Literature Review and Deliverables

We reviewed over 1,200 academic, technical, and selected industry documents across relevant domains (e.g., aviation, defense, healthcare, other transportation) to identify scientific and technical information and data pertaining to the effectiveness, advantages, and limitations of using modern E&DL technologies and approaches for training procedures and pertaining to the design, development, and evaluation of training configurations. These activities allowed us to identify the scientific and technical information and data needed to address our Phase 1 objectives within project deliverables and other information dissemination activities. We periodically submitted interim highlights documents (approx. 13) and submitted semi-annual reports to update the FAA on the project's progress; associated monthly update calls (approx. 15) were conducted jointly with Auburn University to facilitate information sharing. For our Phase 1 deliverables, we developed and submitted to the FAA (Sept 2020) a draft report on the effectiveness of modern E&DL approaches for flightcrew procedures. We then revised and submitted this report to the FAA (Oct 2021); this report discussed the effectiveness, advantages, and limitations of modern E&DL approaches for training flightcrews' procedural knowledge (Sonnenfeld et al, 2021b). These research activities allowed us to share detailed insights from the scientific and technical information and data with our FAA stakeholders.

Throughout the remainder of our period of performance, we supplemented our review of the literature with targeted reviews of topics of relevance to our Phase 2 and Phase 3 objectives. For example, we conducted targeted reviews of the literature on specific flightcrew procedures of interest to the flightcrew training community as identified during our industry discussions (e.g., initial preflight setup/flows, exterior preflight inspection, unreliable airspeed indication, and go-around procedures) to inform our simulation testbed used in Phase 2 and Phase 3 studies. Our initial review of modern training approaches including intelligent tutoring and adaptive training furthered into a targeted review of the implications of artificial intelligence (AI) for VR technologies in training design, development, and assessment practices (see Nguyen et al., 2023) in support of Phase 3 activities. Additionally, informative reviews of regulatory and advisory/guidance materials pertaining to the Advanced Qualification Program (AQP) and Line Operational Simulations (LOS) conducted as part of the RDFSHF-CRM supplemental grant, all helped to inform our Phase 2 and Phase 3 research activities, studies, and the discussion and conclusions of this report.

2.4. Phase 1: Published Work & Outreach

We developed and submitted to the FAA (March 2021) draft abstracts for submission to the International Symposium on Aviation Psychology (ISAP) and the World Aviation Training Summit (WATS). While the WATS submission was not accepted for presentation due to an overwhelming number of submissions, the accepted ISAP submission led to the development and submission to the FAA (April 2021) of an ISAP conference proceedings paper and prerecorded video presentation on findings and suggestions for the air carrier industry on modern training practices in flightcrews' procedural knowledge training. These activities allowed us to disseminate highlights of our literature review activities with the aviation training community. Throughout the remainder of our period of performance, we continued to develop published work which expanded on the topics addressed in these products, as discussed in this report.

Throughout the remainder of our period of performance (March 2021 to September 2023), we continued to develop work products and conduct outreach activities. We developed and submitted to the FAA a series of draft and final versions of five (5) manuscripts for submission to the 2023 Annual Meeting of the Human Factors and Ergonomics Society (HFES) and the 2023 Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), resulting in the acceptance of three (3) proceedings papers and two (2) extended abstracts, as listed in Appendix A. Our outreach activities included participation in an Army Research Laboratory (ARL) Human Research Protection Program (HRPP) visit to UCF/IST (2022) with a presentation on the use of VR in experimental studies related to training, presentation at the UCF AR/VR Innovation Discovery (AVID) symposium (2023), a presentation at the 2023 Human Factors Roundtable (an industry meeting hosted by Atlas Air in Miami in May 2023), and a presentation for the FAA New and Emerging Aviation Technologies (NEAT) series (August 2023).

2.4.1. Overview of Phase 1 Report Findings

Based on our Phase 1 research activities, we determined that flightcrew training providers had begun to offer a number of training solutions delivered using mobile, mixed reality, and VR configurations (see Sonnenfeld et al., 2021b). These solutions were delivered using a range of instructor-led (i.e., facilitated) and non-instructor-led (i.e., non-facilitated) approaches, and there appeared to be a growing acceptance of these emergent technologies among air carriers and among their pilots. We also found that, in practice, various factors related to how the training is designed and implemented (e.g., training time, instructional support, instructional strategies used) may have a larger impact on the effectiveness of training than the type of medium through which training is delivered (e.g., desktop monitor, head-mounted display, mobile device).

Training delivered using immersive technologies appeared to share many of the advantages and limitations of existing simulation training methods delivered through other set-ups (e.g., via CBT, as part of ATDs, etc.). We found VR configurations to be particularly promising for supporting training for the acquisition and practice of procedural knowledge and skill, and we found intelligent tutoring configurations to appear particularly supportive of training of declarative knowledge. That said, there was not yet a deep enough literature to draw firm conclusions. Indeed, one of the main take-aways from Phase 1 was that more information is needed on how the instructional strategies inherent to these configurations (e.g., immersion, interactivity, feedback, scaffolding) support flightcrew training for different tasks and topics. Lastly, we found that contemporary evaluations of training effectiveness should be guided by thorough, systematic methods which account for a breadth of training outcomes and moderating factors. Among these, we have found evidence contrary to the notion that generational differences are a mediating factor in the effectiveness of modern training technologies and approaches: While there may be generational differences in the current pilot population with respect to the degree to which they grew up with computer, gaming, and mobile technologies, we did not find strong support for the notion of a “generational gap”.

As a result of our findings from our review of the literature and supporting discussions, our subsequent research activities in Phase 2 and Phase 3 were focused to provide empirical data to help clarify extant gaps in the literature and in the understanding of the flightcrew training community, particularly with respect to the interactions between immersion and interactivity which may underpin the outcomes of the use of VR technologies over traditional training approaches.

2.5. Discussion: Key Concepts from Phase 1 and Recent Research Findings

Before proceeding to describe the resultant hypotheses and methodologies for our studies conducted under Phase 2 and Phase 3, we recapitulate a few key concepts from our Phase 1 work, supported with background information and recent findings as appropriate. These key concepts include *virtual reality* (VR) in the context of flightcrew training, the concept of *immersion*, and the concept of *interactivity*. An understanding of these concepts is pertinent for contextualizing the hypotheses, methodologies, and results which follow.

2.5.1. Virtual Reality for Flightcrew Training

Virtual Reality (VR) is a term that has variously been used to describe slightly different things: In its essence, virtual reality and virtual simulation can be described as the computer-generated representation of a three-dimensional image or environment that allows individuals to engage with it in a manner that appears genuine or physical. Please note that the broad definition does not specify the display form factor: Both a desktop implementation and one using a head-mounted or another wrap-around display (e.g., a projection in a CAVE virtual environment) would be considered virtual. This is why some authors have distinguished between VR (focus is on VR) and immersive VR (I-VR), where the latter is inherently using an immersive display (cf. McGowin, Fiore, & Oden, 2022).

However, more common is the use of VR as a synonym for I-VR. In immersive VR, which is what we will from here on forward describe with the term VR, the interaction between the trainee and the virtual environment is facilitated through the utilization of specialized electronic equipment, such as a headset equipped with an internal screen or gloves embedded with sensors (Sherman & Craig, 2003; Cummings & Bailenson, 2016; Lawrynczyk, 2018).

VR, especially in its immersive form, is part of an extended reality (XR) spectrum, which differs in the degree to which display elements are real vs. virtual: In pure VR, everything (100%) the user perceives (read: predominantly sees) is virtual and hence computer-generated. In Mixed Reality (MR), some elements are real and seen or otherwise perceived directly (even if mediated by cameras), and what is displayed to the user is blended, with more content being computer generated than real. In Augmented Reality (AR), a few display elements are computer generated and are overlayed or otherwise fused with the predominant portion of what is observed by the user, being real. Together, VR, MR, AR, make up the XR spectrum.

The aviation industry has begun to embrace the potential of innovative XR training techniques and tools for flight crew training, such as immersive VR (Risukhin, 2016). Aviation instruction plays a vital part in the aviation industry as it guarantees that pilots and other aviation professionals acquire the necessary information, skills, and competencies to perform their roles safely and effectively (Jentsch & Curtis, 2017). VR shows considerable promise in its application to flight crew training, but the technology is not without its challenges, particularly in relation to how interactivity and immersion affect training fidelity, as well as specific technology limitations, such as the potential for VR-/cybersickness. A comprehensive exploration of VR’s efficacy in both short-term and long-term pilot training is still lacking (Oberhauser et al., 2018; Risukhin, 2016). Thus, in addition to assessing the overall effectiveness of employing VR technology, the optimal methods for its utilization continue to necessitate additional investigation (Walters & Walton, 2021).

Like other simulation approaches and training devices, those using VR technology could enable pilots to repetitively engage in training objectives or scenarios, including some that are deemed unsafe to be conducted in person (Orlansky et al., 1994). VR simulators can reduce costs by minimizing the necessity of utilizing actual aircraft for training purposes, while also enabling the repetition of training tasks in a rapid and effective manner. The efficacy of flight simulators is contingent upon the pilot training objective (Roscoe & Bergman, 1980) and the specific phase of training. According to Povenmire and Roscoe (1973), novice pilots may derive greater advantages from utilizing the simulator compared to their more experienced counterparts (Reyes et al., 2022; Povenmire & Roscoe,1973). The findings of several studies conducted on the use of virtual reality (VR) in aviation indicate that this technology has the potential to enhance pilot training by improving the quality of instruction or the way content is delivered (Lawrynczyk, 2018; Lewis & Livingston, 2018; Schaffernak et al., 2020).

A useful three-dimensional framework to conceptualize, classify, and compare XR solutions considers *immersion* and *interactivity*, as originally proposed for VR by, among others, Slater and Wilbur (1997), and adds *realism* as a third dimension (Figure 1). In this context, immersion refers to the degree to which the XR solution provides a sense of vivid presence in the target environment; interactivity to the way the experience is changed from passive to active; and realism describes how the XR experience approximates the existing reality of the target environment. We discuss these next, in turn.

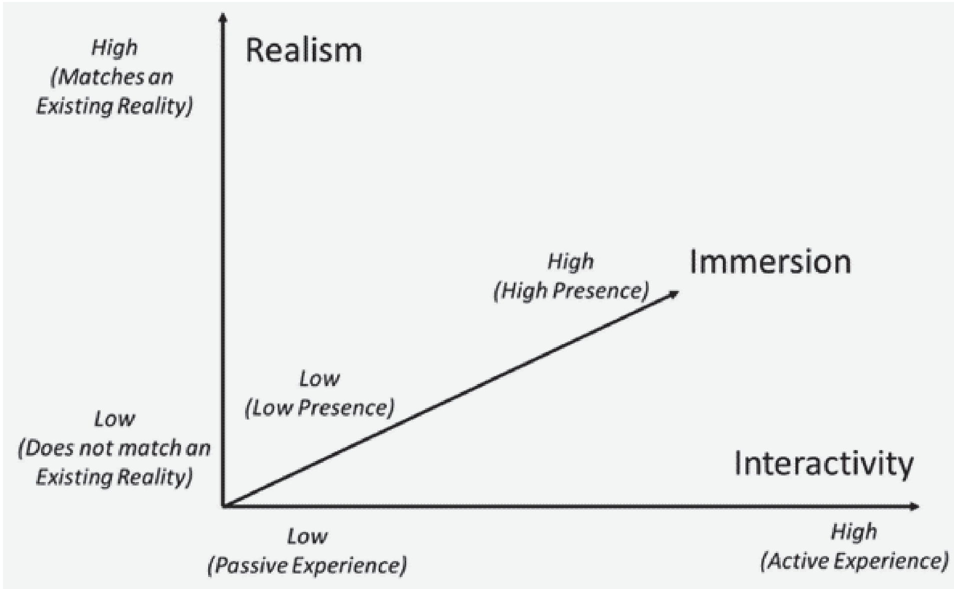


Figure 1. Three-dimensional framework to conceptualize, classify, and compare VR solutions

Note that the three dimensions are near orthogonal to one another: A fantasy/sci-fi VR game, for example, may be highly immersive (creating a vivid sense of presence) and highly interactive (providing lots of opportunities to interact with and affect the environment), but it may be completely lacking in realism (e.g., ignore the physics of gravity, force, etc.). Conversely, a telepresence VR experience, such as one to visit a tourist attraction like Times Square in New York City, may provide full 360-degree video and audio immersion and stunning realism in 8K UHD, but it may lack any opportunity for interactions. And a nuclear power station control room systems trainer implemented on a tablet display may fall short on immersion, but it may be highly interactive and be realistic in its depiction of the control room displays and their behavior.

2.5.2. Immersion

Immersion, in the context of virtual reality (VR), refers to the degree to which a user feels fully absorbed and engaged in a computer-generated environment, to the point where they perceive themselves as being physically present within that environment. It involves the sensory and perceptual factors that make the virtual experience seem real and convincing (Slater & Wilbur, 1997; Lombard & Ditton, 1997; Slater, 2003). Its meaning is characterized as "allowing users to complete their entry into the environment presented on the screen," which in turn gives them the impression that they are not only observers but also an integral part of the virtual world. Immersion is determined by several factors, including the number of sensory modalities used, the congruency of those modalities (e.g., visuals matching the user's head motion), the field of view, the vividness of the experience (e.g., frame rate, resolution), the user's ability to influence events, and the plot that develops. For a review of constructs affecting immersion, see Slater & Wilbur, 1997; Miller & Bugnariu, 2016; and Sanchez-Vives & Slater, 2005.

The Cognitive Affective Model of Immersive Learning (CAMIL) proposes that two technological aspects of virtual reality (VR), interactivity and immersion, influence a number of cognitive and affective variables that may facilitate or impede learning in virtual learning environments (VLE). Immersion is not a unique characteristic in and of itself; rather, it is the outcome of representational integrity and interactivity (Dalgarno and Lee, 2010; Johnson-Glenberg, 2018; Makransky & Petersen, 2021; Peterson, et al., 2022). The primary differentiating element between a virtual reality (VR) learning session accessed by a head-mounted display (HMD) as opposed to a VR session accessed through a desktop computer, is the level of immersion experienced. Immersion is a quantifiable metric that assesses the level of realism and sensory engagement provided by a system, as well as its ability to minimize external distractions (Cummings & Bailenson, 2016).

2.5.3. Interactivity

Interactivity pertains to the degree of autonomy granted to the user in influencing the learning experience typically facilitated by handheld controllers and a virtual embodiment (Makransky & Petersen, 2021). Makransky et al. (2020a) suggest that VR is different from more standard multimedia lessons like videos or PowerPoint because of the level of interaction that is possible. Interactivity is a pivotal element in virtual reality (VR), enabling users to actively engage with digital environments and fostering a profound sense of immersion. Through a combination of hardware and software, VR allows users to interact with and manipulate the virtual world in real-time, creating an unprecedented level of agency and presence (Slater & Wilbur, 1997; Seymour et al., 2002; Maher et al., 2000; Dalgarno & Lee, 2010).

2.5.4. Realism

Realism describes the degree to which the simulation environment replicate reality; it is often described also as “fidelity”. The realism of the physical representation of an environment (e.g., what it looks like) is subsumed under physical fidelity, and maps closely onto immersion. That is, immersive displays that accurately replicate the metric space and visual appearance of an environment have high physical fidelity (at least as far as the visual representation is concerned). In contrast, realism of the functional representation of an environment (e.g., how it works) is termed functional fidelity, and this relates to the degree of interactivity. Consequently, highly interactive displays, that correctly represent the actions of users and the consequences of their actions, is said to have high functional fidelity.

A third dimension of fidelity is what has been referred to as “psychological fidelity”. It refers to the degree to which a system or environment stimulates the targeted psychological constructs of the environment, which could be sensory, perceptual, cognitive, or affective. Some training researchers (beginning with Osgood) argue that transfer of training is facilitated by high psychological fidelity, and that it is more important than either physical or functional fidelity. However, because psychological fidelity is harder to verify and validate (it inherently depends on the human), it is typically easier to specify, verify, and validate the physical and functional fidelity of the simulation or training environment. If both are high, it is assumed that psychological fidelity is high as well, whereas when one or the other, or both, are not, this assumption is at best tenuous.

2.6. Overview of Research Questions

Having recognized the potential of VR simulations, our Phase 2 and Phase 3 research sought to investigate immersion and interactivity using various simulation environments and instructional elements. Our aim was to determine the degree to which interactivity might compensate for immersion, especially for tasks that have different levels of involved motor actions. To obtain answers to these questions, we compared different instructional elements in two implementations (i.e., a more immersive (VR using HMD) versus a less immersive (PC with monitor/Tablet). Simultaneously, we measured instances and severity of cybersickness, and we sought feedback from study participants regarding their reactions, perceptions, and opinions of the effectiveness and user-friendliness of the simulations. Moreover, we asked the learners to give us their perceived instrumentality and usability of the simulations. For example, we asked the learner to provide their thoughts on the training that they received, what they liked or disliked about the training, and around what time or point did they feel familiar or comfortable with the VR system.

For each Phase, we formulated a set of research questions, shown in the next two subsections.

2.6.1. Phase 2 Study Research Questions

For Phase 2, the main research questions were the interaction, usability, the occurrence of cyber sickness, and learning experiences for task novices with a comparative sample (over 100 participants; college students) and an achievable task (exterior preflight inspection). Participants went through a training for an external preflight inspection of a transport-category aircraft; and we compared immersive (VR with HMD) and less immersive (PC with monitor), as well as highly interactive (active exploration: animations, highlighting, dynamic comparisons and others, in addition to static comparison, text) and less interactive (limited exploration: static comparison, text) instructional elements.

- RQ1: How does the immersiveness of the system, the interactivity of instruction, and the interaction between these constructs influence variance in the effectiveness of training for flightcrews’ procedures?
- RQ2: How does the level of interaction differ between highly immersive simulations (VR) and less immersive simulations (PC) for novices performing an exterior preflight inspection task?
- RQ3: What is the impact of highly immersive simulations (VR) versus less immersive simulations (PC) on learning experience for novices?
- RQ4: What is the perceived usability of highly immersive simulations VR) versus less immersive simulations (PC) for novices?

2.6.2. Phase 3 Study Research Questions

For Phase 3, which had access to a smaller, but highly qualified sample of active air carrier pilots at a major US airline, the focus was on questions related to perceived usability, instrumentality, acceptability, and cybersickness. The air carrier pilots each interacted with two different training implementations, a trainer for external preflight inspections and one for initial preflight setup flows and compared immersive (VR with HMD) and less immersive (PC with monitor/tablet) implementations of the simulations.

- RQ1: How are Part 121 air carrier pilots' affective responses (reactions, perceptions, and opinions) toward training approaches influenced by certain system qualities (e.g., level of immersion) across different procedures (e.g., preflight inspection, initial preflight setup)?
- RQ2: What are the practical considerations and challenges associated with implementing highly immersive and less immersive simulations in the training pipeline?
- RQ3: To what extent do highly immersive simulations (VR) and less immersive simulations (PC/tablet) differ in terms of usability when employed as a training tool for two selected tasks, exterior preflight inspection and initial preflight setup/flows in a generic multiengine aircraft flight deck?
- RQ4: What are the optimal ways to integrate virtual reality training into existing flight crew training programs to maximize its effectiveness and complement traditional training methods like the paper tigers and flight simulators?

3.0. Phase 2 Empirical Study (In-house Laboratory Study with College Students)

Modern approaches to flightcrew training may vary in effectiveness due to (a) various design differences, and (b) for different types of tasks. Through our current line of research, we investigated the effectiveness of various modern training configurations using different instructional strategies, across a series of scenarios involving procedures associated with flight operations. This study focused on capturing individuals' learning gains following a mock Preflight Exterior Inspection task experienced through a simulation, Flight Procedures Experimental Training (FlightPET).

The primary purpose of this study was to develop an understanding of how students acquire the knowledge and skills to perform procedures during flight operations (here, preflight exterior aircraft inspection). One of our key interests was to determine the differences in effectiveness between training for these procedures using personal computer (PC)-based and virtual reality (VR)-based simulation devices. We were also interested in determining how different combinations of instructional strategies impact the effectiveness of training using these devices [i.e., passive strategies (e.g., static text & media) versus active strategies (e.g., dynamic interactions, object enrichment, cues & clues)]. Based on prior review of the literature, we were interested in evaluating training effectiveness using multiple types of indicators (i.e., affective effectiveness, behavioral effectiveness, cognitive effectiveness, process effectiveness), while accounting for various pre-training and training factors (e.g., individual differences, prior knowledge, motion sickness, time variables).

This study used undergraduate university students to investigate how students acquire the knowledge and skills to perform procedures during flight operations. Participants were assigned to one of four conditions: PC-based simulation, with active or passive instructional strategies; and VR-based simulation, with active or passive instructional strategies. A more detailed description of the method is presented in this section.

3.1. Hypotheses

- *Hypothesis 1.* There will be an interaction effect such that the low immersion x low interactivity condition and high immersion x high interactivity conditions result in equivalently high learning gains, but with trade-offs in various affective (e.g., presence, usability, satisfaction), cognitive (e.g., workload, self-regulation), and process (e.g., training time, simulation sickness) indicators.
- *Hypothesis 2.* With respect to the preflight exterior aircraft scenario, the active instructional strategy combination will demonstrate evidence of higher training effectiveness.
- *Hypothesis 3.* There will be no significant differences in training effectiveness overall between the PC and VR configurations.

3.2. Method

3.2.1. Participants

The study sample for Phase 2 was made up of *N* = 102 undergraduate students recruited from the University of Central Florida (UCF) who participated in this study for course credit. As the goal of this study was to understand how immersion and interactivity may impact novices' learning experience and resultant acquisition of flightcrew-relevant knowledge and skills, college students represented a suitable population. None of these undergraduate students reported prior experience as pilots. Of the 102 participants, 52 reported as male (51%) and 49 as female (48%) (1 other). The sample included 90 participants aged 18-20 (88.2%), 11 participants aged 21-30 (10.8%), with 1 participant in the 31-40 category (1%). Thus, more than 95% of the sample were representative of the next generation of air carrier pilots.

Participants were randomly assigned to the four study conditions. No significant differences were found between conditions for the demographic characteristics collected for the study (e.g., dominant handedness, corrective vision, biological sex). Students were recruited using a UCF platform and received course credit for their participation in this study.

3.2.2. Design

In order to examine the relationship between (1) immersion and (2) interactivity on training effectiveness, with a focus on declarative knowledge and procedural skill outcomes in a flightcrew training context (e.g., exterior preflight inspection of a generic multiengine aircraft), this study used a 2 (Immersion; Lower vs. Higher) X 2 (Interactivity; Lower vs. Higher) between-subject factorial design (see Table 1).

Table 1

Phase 2 Study Design

		Interactivity	
		Low	High
Immersion	Low	**PC Simulation with 2D** Display and Peripherals; Passive Instructional Strategy Set	**PC Simulation with 2D** Display and Peripherals; Active Instructional Strategy Set
	High	**VR Simulation with 3D** Display and Peripherals; Passive Instructional Strategy Set	**VR Simulation with 3D** Display and Peripherals; Active Instructional Strategy Set

3.2.3. Apparatus

The study was administered on an Alienware desktop computer and HTC Vive Focus 3 virtual reality (VR) headset for the portion of the experiment (see Figure 2). The Flightcrew Procedures Experimental Training (FlightPET) simulation (preflight inspection) developed by UCF as part of this project (Sonnenfeld et al., 2023) was utilized in both PC-based and VR-based conditions (see Figure 3 and Figure 4). FlightPET was developed to support research on modern training practices in the air carrier industry and was designed to help identify which combinations of display technologies and training approaches should be used for specific flightcrew-relevant topics, in what contexts, and how to leverage these configurations to create outcomes comparable to conventional training methods. It can be implemented using virtual reality (VR) or conventional desktop displays. Technical details on FlightPET, a description of its features, and a discussion of its requirements, iterative design, testing, and potential use, can be found in Sonnenfeld et al. (2023) - see also Appendix A.

Figure 2. VR Configuration & Controls (HTC Vive Focus 3)

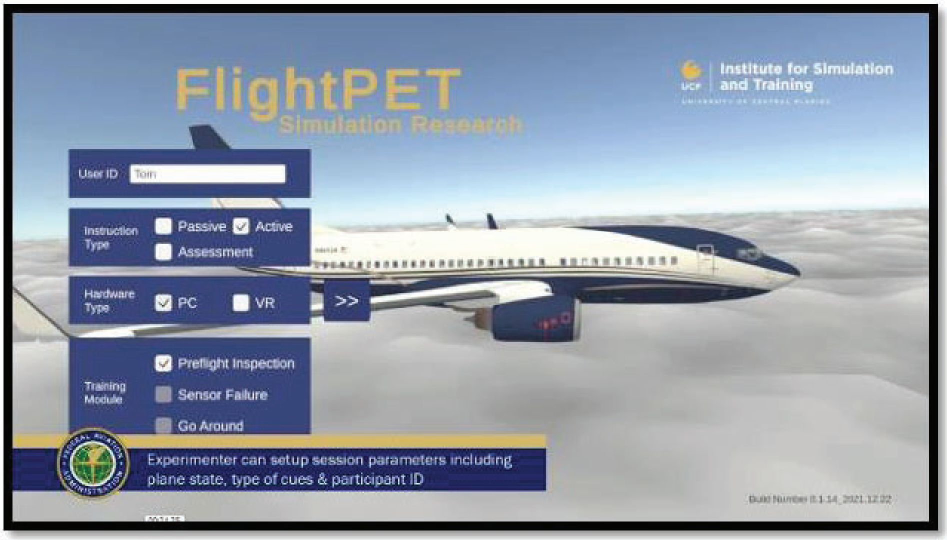


Figure 3. UCF FlightPET experimentation setup menu

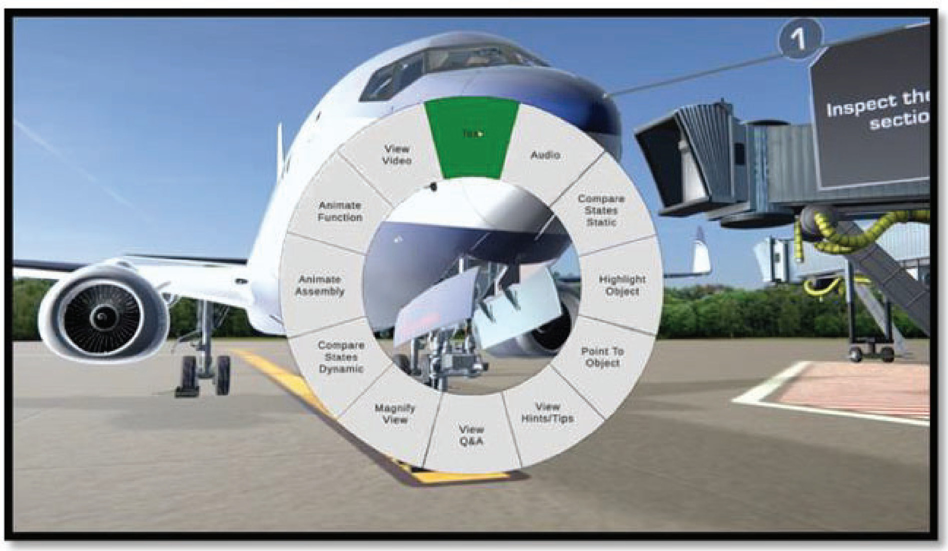


Figure 4. FlightPET preflight exterior inspection training environment

3.2.4. Materials & Measures

3.2.4.1 *Informed Consent Form.* Participants were greeted at the research facility and presented with an informed consent form. This form explained to the participants the purpose of the study and any potential risks associated with participating in the study. In addition, participants were told that their involvement in the study was strictly voluntary and that they could decline or withdraw from participation at any point of the study.

3.2.4.2 *Study Conditions.* Participants were randomly assigned to one of four study conditions, a) PC-based Simulation with Passive Instructional Strategies, (b) PC-based Simulation with Active Instructional Strategies, (c) VR-based Simulation with Passive Instructional Strategies, or (d) VR-based Simulation with Active Instructional Strategies).

3.2.4.2.1 Condition A1: PC-based Simulation with Passive Instructional Strategies. Participants assigned to the "PC - Passive" condition were trained only with passive instructional strategies (e.g., static text and media) on the FlightPET desktop training simulation testbed.

3.2.4.2.2 Condition A2: PC-based Simulation with Active Instructional Strategies. Participants assigned to the "PC - Active" condition were exposed to a training environment with active instructional strategies (e.g., dynamic interactions, object enrichment, cues, and clues) on the FlightPET desktop training simulation testbed.

3.2.4.2.3 Condition B1: VR-based Simulation with Passive Instructional Strategies. Participants assigned to the "VR - Passive" condition were trained only with passive instructional strategies (e.g., static text and media) on the FlightPET virtual reality training simulation testbed.

3.2.4.2.4 Condition B2: VR-based Simulation with Active Instructional Strategies. Participants assigned to the "VR - Active" condition were exposed to a training environment with active instructional strategies (e.g., dynamic interactions, object enrichment, cues, and clues) on the FlightPET virtual reality *training* simulation testbed.

3.2.4.3 *Pre-Testing Measures.* Before engaging in their assigned experimental condition, participants were asked to complete several study measures on Qualtrics. These measures consisted of questions regarding participants age range, aviation experience, and individual difference factors including personality tendencies, experience using technologies including computer-based flight simulations and virtual reality head-mounted displays, and prior domain knowledge. Additionally, participants completed a color vision test as well as a spatial visualization test. Both tests were used for data screening purposes only and were not used to determine participants' eligibility/ineligibility. The Corsi block-tapping test (Corsi, 1972; see also Arce & McMullen, 2021) was used to measure visuospatial working memory (VSWM; Kessels et al., 2008; Vandierendonck et al., 2004). This test essentially required the subject to observe a sequence of highlighted blocks and then to repeat the sequence, beginning with a small number of blocks and gradually increasing in length up to nine blocks. Collectively, these pre- and post-surveys were used to account for potential confounds and to better understand individual differences in training effectiveness.

3.2.4.4 *Post-Testing Measures.* After engaging in their assigned experimental condition, participants were asked to complete a series of assessments on Qualtrics (e.g., performance of preflight inspection task in FlightPET simulation, knowledge assessment, transfer assessment, collection of post-training survey responses, and obtaining feedback from participants). Moreover, affective (e.g., presence, usability, satisfaction), behavioral (e.g., task performance), cognitive (e.g., learning gains, cognitive workload, self-regulation), and process (e.g., time, simulation sickness) constructs were used as indicators of training effectiveness.

3.2.5. Procedure

After arrival, participants were randomly assigned to one of the four conditions described above. First, participants were briefed on the purpose and procedures of this study. Second, participants were asked to respond to some general surveys which asks about information such as participants age range, aviation experience, and individual difference factors including personality tendencies, experience using technologies including computer-based flight simulations and virtual reality head-mounted displays, prior domain knowledge. Additionally, participants were asked to complete a color vision test as well as a spatial visualization test. After completing the initial surveys, participants received general training on the controls that they would need to be familiar with to successfully engage with the mock tasks within our flight simulation software used for this study (either a standard desktop computer, mouse, and keyboard; or a virtual reality head-mounted display and hand-held controllers). Afterwards, participants were tasked with performing a virtual preflight exterior aircraft inspection task without training aids. This task would be followed by a competency test. The inspection task and the competency test are assessments of the simulation and were administered to all participants. After these assessments participants were asked to complete related to motion sickness, perceived workload, user experience, and other reactions to the training they received. In addition, they were also asked to provide feedback about their experience, understanding of the task, and the training they received. Finally, participants were debriefed on the purpose of the study and protocols.

3.3. Results

3.3.1. Data Checks for Pre-Study Measures and Random Assignments

We began our analyses by comparing the four participant groups, to ensure equivalency across important background variables and biodata. Initial data checks showed that the four groups were comparable across the vast majority of background variables. Using one-way ANOVAs to compare the four groups, there were no substantial nor any significant ($p < .05$, two-tailed) differences in subgroup averages for age group, TIPI scores on extraversion, agreeableness, conscientiousness, emotional stability, and openness to experience. Likewise, there were no significant differences in need for cognition (NFC-6, global). There were also no significant differences by study group in pre-experiment scores on the knowledge test (baseline knowledge).

With respect to the self-reported visually-induced motion sickness susceptibility (VIMSS), there was no significant difference among the four subgroups when compared directly, $F\ (3,98) = 1.518$, $p = .2189$. However, when we compared the means across the four groups, we noted that participants in the two low-interactivity conditions seemed to each have higher VIMSS mean scores than the respective high-interactivity groups (see Figure 5 below), so we followed the one-way ANOVA up with a two-way ANOVA.

Table 2

ANOVA Table for the Self-Reported Motion Sickness Susceptibility across Groups in the Phase 2 Study

Tests of Between-Subjects Effects

Dependent Variable: msstot

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	35.162[a]	3	11.721	1.502	.219
Intercept	700.762	1	700.762	89.795	<.001
immerse	.149	1	.149	.019	.890
interact	34.913	1	34.913	4.474	.037
immerse * interact	1.099	1	1.099	.141	.708
Error	764.798	98	7.804		
Total	1536.000	102			
Corrected Total	799.961	101			

a. R Squared = .044 (Adjusted R Squared = .015)

This analysis indicated that when combined, participants randomly assigned to the two low-interactivity groups reported, on average, significantly higher VIMSS scores (M = 3.3) than participants in the high-interactivity conditions (M = 2.1), $F(1, 98) = 4.474$, $p = 0.037$. There was, however, neither a main effect for immersion ($p = 0.890$), nor an interaction effect ($p = 0.708$). The observed effect size $\eta 2$ for the relationship of VIMSS by interactivity was small, $\eta 2$=0.044 (adjusted, 0.015), however, and so we noted the difference for subsequent analyses but did not correct for it.

Note: A = VR; B = PC; 1 = Passive; 2 = Active
A1: Baseline text and static images, presented in VR
A2: Baseline text and static images, plus dynamic comparisons and animations, presented in VR
B1: Baseline text and static images, presented on PC monitor
B2: Baseline text and static images, plus dynamic comparisons and animations, presented on PC monitor

Figure 5. Plot of self-reported motion sickness susceptibility across study groups in Phase 2.

3.3.2. Study 2: Post-Participation Measures

3.3.2.1. Post-Training (T1) Cybersickness Results

A two-way ANOVA was conducted to analyze the main effects and interaction of the level of immersion (high VR, low PC) and interactivity (more active, passive) on cybersickness, as measured by the SSQ Total Scores after training (T1). As detailed in Table 3, we observed a main effect of immersion on cybersickness scores, $F(1, 98) = 14.015$, $p < .001$. The observed effect size η^2 for the effect of immersion was medium, $\eta 2 = 0.12$. There was not a main effect for interactivity on cybersickness, nor a statistically significant interaction between immersion and interactivity.

Table 3

ANOVA Table for the Total SSQ Scores after Training (T1) in the Phase 2 Study

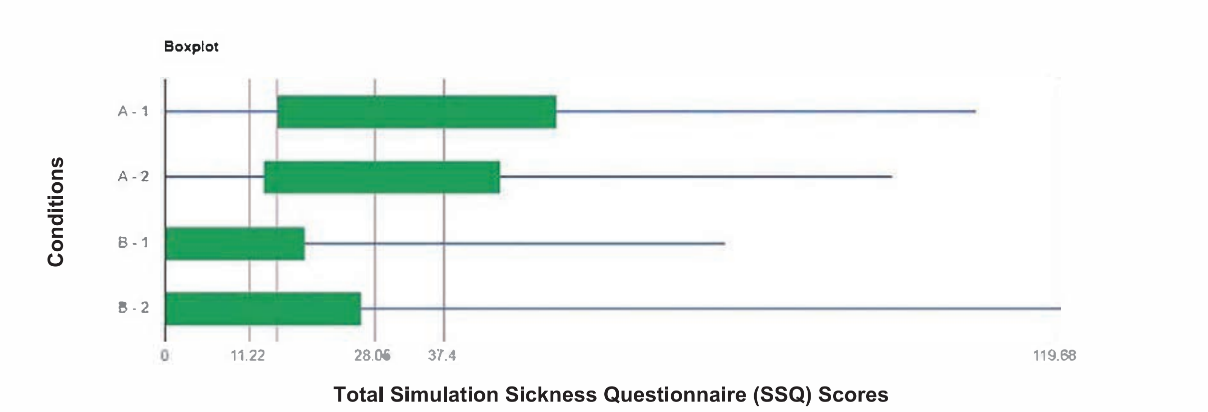
Tests of Between-Subjects Effects

Dependent Variable: t1ssqtot

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7744.431[a]	3	2581.477	4.757	.004
Intercept	66838.250	1	66838.250	123.173	<.001
immerse	7605.201	1	7605.201	14.015	<.001
interact	21.708	1	21.708	.040	.842
immerse * interact	211.816	1	211.816	.390	.534
Error	53178.424	98	542.637		
Total	121020.715	102			
Corrected Total	60922.855	101			

a. R Squared = .127 (Adjusted R Squared = .100)

As shown in Figure 6, participants' average total SSQ scores in the VR Active ($M = 33.29$) and VR Passive ($M = 37.20$) conditions showed that these participants experienced significantly and substantively more severe symptoms of cybersickness, than participants in the PC Active ($M = 18.47$) and PC Passive ($M = 16.46$) conditions, and the degree of variance was in part statistically attributable to the immersion provided by the modality of the system modality ($\eta^2 = 12\%$). Participants' reported experience of these symptoms nearly doubled between the high immersion (VR) and low immersion (PC) conditions.



Note: A = VR; B = PC; 1 = Passive; 2 = Active
A1: Baseline text and static images, presented in VR
A2: Baseline text and static images, plus dynamic comparisons and animations, presented in VR
B1: Baseline text and static images, presented on PC monitor
B2: Baseline text and static images, plus dynamic comparisons and animations, presented on PC monitor.

Figure 6. Plot of self-reported Total SSQ scores after Training (T1), across Phase 2 study groups

3.3.2.2. Post-Performance Task (T2) Cybersickness Results

A two-way ANOVA on cybersickness severity as measured by the Total SSQ scores after the performance task (T2) was conducted to analyze the main effects and interaction of the level immersion (high, low) and interactivity (active, passive). As shown in Table 4, the analysis indicated that there was a statistically significant effect of immersion on cybersickness, $F(1, 97) = 12.626$, $p < .001$, medium effect size $\eta 2 = 0.125$. However, there was neither a statistically significant effect of interactivity, nor of the interaction between immersion and interactivity.

Table 4

ANOVA Table for the Total SSQ Scores after the Performance Evaluation (T2) in the Phase 2 Study

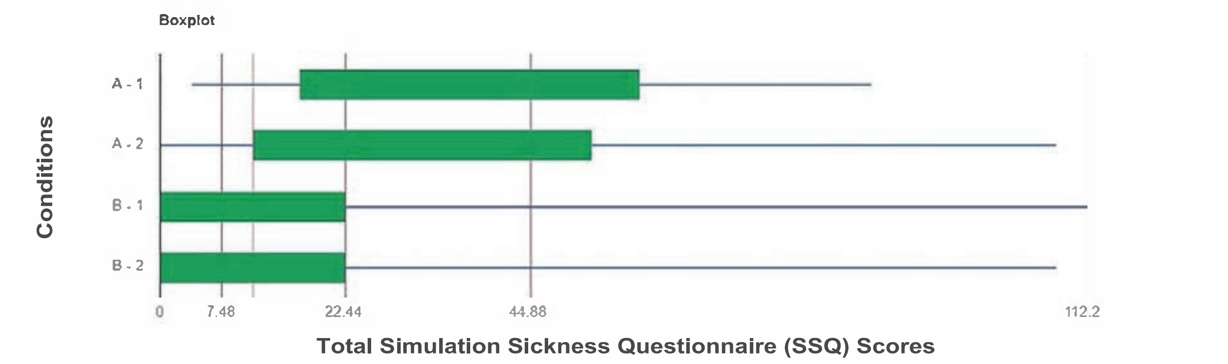
Tests of Between-Subjects Effects

Dependent Variable: TSSQ2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	9397.688[a]	3	3132.563	4.629	.005
Intercept	67624.105	1	67624.105	99.918	<.001
IMM	8545.399	1	8545.399	12.626	<.001
INT	797.327	1	797.327	1.178	.280
IMM * INT	63.814	1	63.814	.094	.759
Error	65649.248	97	676.796		
Total	134826.476	101			
Corrected Total	75046.936	100			

a. R Squared = .125 (Adjusted R Squared = .098)

As shown in Figure 7, participants' total SSQ scores in the VR Active (*M* = 32.35) and VR Passive (*M* = 39.76) conditions showed that these participants experienced significantly and substantively more acute symptoms of cybersickness than participants in the PC Active (*M*=15.07) and PC Passive (*M* = 19.22) conditions; in fact, approximately double the level and, in all cases, above 20 for the VR, and below 20 for the PC versions. Similar to the T1 data analysis, participants' reported experience of these symptoms nearly doubled between the high immersion (VR) and low immersion (PC) conditions.



Note: A = VR; B = PC; 1 = Passive; 2 = Active
A1: Baseline text and static images, presented in VR
A2: Baseline text and static images, plus dynamic comparisons and animations, presented in VR
B1: Baseline text and static images, presented on PC monitor
B2: Baseline text and static images, plus dynamic comparisons and animations, presented on PC monitor

Figure 7. Plot of self-reported Total SSQ scores after evaluation (T2), across Phase 2 study groups

3.3.2.3. Post-Training (T1) Knowledge Assessment Results

Our next set of analyses focused on the post-training knowledge assessment results. We conducted two sets of analyses: We analyzed the absolute post-test scores, and then calculated gain scores between the pre-participation assessment and the post-training knowledge assessment to determine how much learning had taken place. We also conducted ANCOVA [using pretest scores as the covariate] and mixed-model analyses, all of which showed the same pattern of results, however. In the following, we report first, the result of the post-test scores only analysis, and then of the gain score analysis. For completeness, we also report the results of the mixed model analysis.

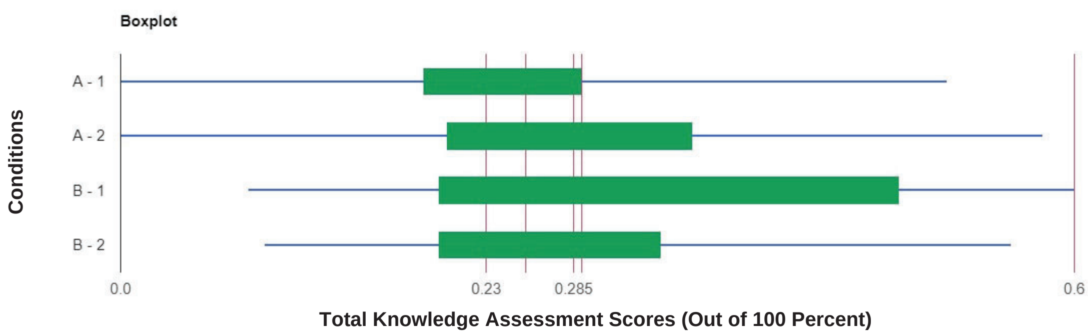
First, a two-way ANOVA was conducted to analyze the main effects and interaction of the level immersion (high, low) and interactivity (active, passive) on total knowledge assessment scores. As detailed in Table 5, the analysis indicated that there was not a statistically significant (a=0.05) effect on total knowledge assessment scores due to the interaction between immersion and interactivity [$F(1,98) = 2.61$, $p = 0.1093$]. Main effects analysis showed that immersion had no significant effect on total knowledge scores (p = 0.1353). There was not a main effect for interactivity on total knowledge assessment scores (p = 0.6901).

Table 5

ANOVA Table for the Post-Training (T1) Knowledge Assessment Results Phase 2 Study

Source	DF	Sum of Square (SS)	Mean Square (MS)	F Statistic (df_1, df_2)	P-value
Factor A - rows (immerse)	1	0.04018	0.04018	2.2684 (1,98)	0.1353
Factor B - columns (interact)	1	0.002832	0.002832	0.1599 (1,98)	0.6901
Interaction AB	1	0.04626	0.04626	2.6115 (1,98)	0.1093
Error	98	1.736	0.01771		
Total	**101**	**1.8252**	**0.01807**		

As shown in Figure 8, participants' total knowledge assessment scores after training in the VR Active (M = 0.28), VR Passive (M = 0.24), PC Active (M = 0.28), and PC Passive (M = 0.33) conditions did not differ statistically or substantially. Neither immersion nor interactivity seemed to have had an effect on total knowledge assessment scores.



Note: A = VR; B = PC; 1 = Passive; 2 = Active
A1: Baseline text and static images, presented in VR
A2: Baseline text and static images, plus dynamic comparisons and animations, presented in VR
B1: Baseline text and static images, presented on PC monitor
B2: Baseline text and static images, plus dynamic comparisons and animations, presented on PC monitor

Figure 8. Plot of the knowledge assessment scores after training (T1) across Phase 2 study groups (e.g., 0.25 = 25 percentage points).

3.3.2.4. Baseline-to-post-Training Knowledge Score Differences (Gain Scores)

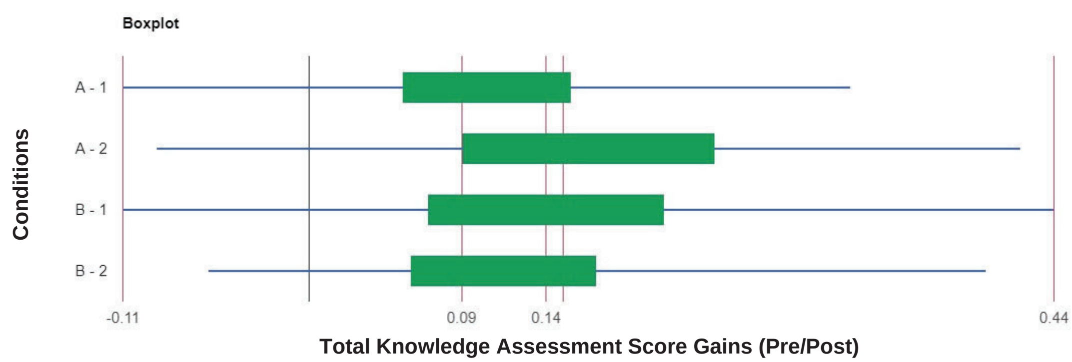
A two-way ANOVA was conducted to analyze the main effects and interaction of the level immersion (high, low) and interactivity (active, passive) on baseline-to-total knowledge differences (knowledge gains). As detailed in Table 6, the analysis indicated that there was not significant effect of immersion (supporting hypothesis H3) nor of interactivity (contrary to hypothesis H2). However, the interaction of immersion and interactivity approached statistical significance, F (1,98) = 3.9327, p = 0.05015.

Table 6

ANOVA Table for the Knowledge Gain Scores (T1 – Baseline) after T1 in the Phase 2 Study

Source	DF	Sum of Square (SS)	Mean Square (MS)	F Statistic (df_1, df_2)	P-value
Factor A - rows (immerse)	1	0.003296	0.003296	0.2864 (1,98)	0.5938
Factor B - columns (interact)	1	0.0002428	0.0002428	0.0211 (1,98)	0.8848
Interaction AB	1	0.04526	0.04526	3.9327 (1,98)	0.05015
Error	98	1.1279	0.01151		
Total	**101**	**1.1767**	**0.01165**		

As shown in Figure 9, participants' knowledge gains (baseline-to-total knowledge assessment scores) were uniformly greater than 0 and in the range of 10 to 16 percentage points, or roughly doubling the scores from pre- to post-test. Specifically, the averages among participants in the VR Active (*M* = 0.16 [16 percentage points gain over baseline]) and PC Passive (*M* = 0.16) conditions were slightly higher than in the VR Passive (*M* = 0.10) and PC Active (*M* = 0.13) conditions, albeit not statistically significantly different. This pattern was different from the one we had expected in Hypothesis 1: Instead of increased interactivity making up for loss of immersion, here, both the most immersive and most interactive (VR Active) and the least immersive and least interactive (PC Passive) conditions showed the greatest gains, albeit not significantly higher than their respective counterparts in terms of interactivity.



Note: A = VR; B = PC; 1 = Passive; 2 = Active
A1: Baseline text and static images, presented in VR
A2: Baseline text and static images, plus dynamic comparisons and animations, presented in VR
B1: Baseline text and static images, presented on PC monitor
B2: Baseline text and static images, plus dynamic comparisons and animations, presented on PC monitor

Figure 9. Plot of the baseline-to-T1 knowledge score gains by Phase 2 study group (e.g., 0.10 = 10 percentage points).

A follow-up test using a 2 (immersion) x 2 (interactivity) x 2 (time) mixed-model ANOVA on the pre-training and post-training knowledge assessment stores confirmed that in all four conditions, the knowledge scores increased significantly, both overall $F(1, 98) = 153.27$, $p < .001$, *Partial Eta*^{2}

= .610, and within each of the four groups (using simple effects test of time in each of the four conditions). However, when comparing the size of the gains over time across study conditions, the only significant effect was the simple effect of immersion (VR vs PC) at low interactivity (p = .048): Under conditions low interactivity, the VR group gained significantly less knowledge than the PC group.

3.3.2.5. Post-Performance Task (T2) System Usability Score (SUS) Results

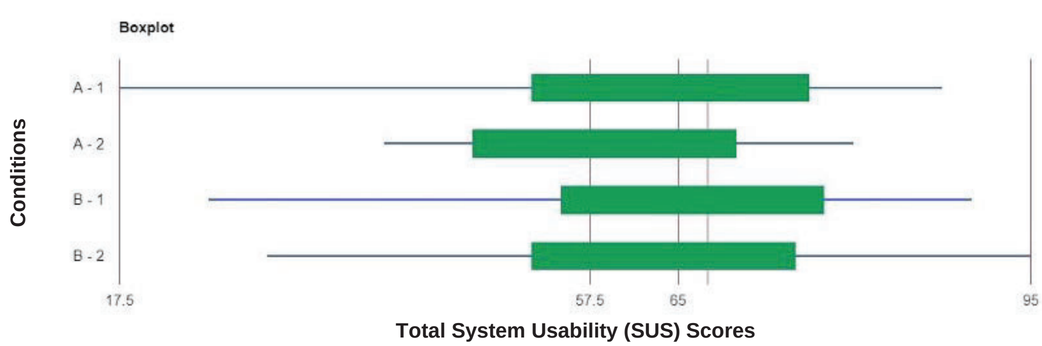
A two-way ANOVA was conducted to analyze the main effects and interaction of the level of immersion and interactivity on usability rating scores as measured by the SUS at T2. As detailed in Table 7, the analysis indicated that there were no statistically significant differences in usability ratings between conditions.

Table 7

ANOVA Table for System Usability Scores after T2 in the Phase 2 Study

Source	DF	Sum of Square (SS)	Mean Square (MS)	F Statistic (df_1, df_2)	P-value
Factor A - rows (immerse)	1	353.5124	353.5124	1.2549 (1,98)	0.2654
Factor B - columns (interact)	1	7.6927	7.6927	0.02731 (1,98)	0.8691
Interaction AB	1	35.5555	35.5555	0.1262 (1,98)	0.7232
Error	98	27607.9575	281.7139		
Total	**101**	**28004.7181**	**277.2744**		

The average usability scores in the VR Active (M = 59.25), VR Passive (M = 61.32), PC Active (M = 69.24), and PC Passive (M = 63.92) conditions had no statistically or substantially notable differences based on immersion or interactivity. However, PC conditions were rated slightly higher than the VR conditions.



Note: A = VR; B = PC; 1 = Passive; 2 = Active
A1: Baseline text and static images, presented in VR
A2: Baseline text and static images, plus dynamic comparisons and animations, presented in VR
B1: Baseline text and static images, presented on PC monitor
B2: Baseline text and static images, plus dynamic comparisons and animations, presented on PC monitor

Figure 10. Plot of the post-assessment (T2) system usability (SUS) scores across Phase 2 study groups

Similar to the SUS scores, analysis of assessments of workload using the subset of the NASA TLX scales used in this study, showed no significant differences in average workload ratings (aggregated) either due to immersion or interactivity.

4.0. Phase 3 Empirical Study (Field Study at Training Center with Air Carrier Pilots)

Modern approaches to flightcrew training may vary in effectiveness due to (a) various design differences, and (b) for different types of tasks. In Phase 3, we investigated the effectiveness of various modern training configurations using different instructional strategies, across a series of tasks involving procedures associated with air carrier operations (e.g., preflight inspection, preflight setup). Our study focused on developing an understanding of how 14 CFR Part 121 air carrier pilots acquire the knowledge and skills to perform procedures relevant to flight operations, by capturing participants' perceptions, reactions, and opinions of different training approaches for flightcrew-relevant tasks. Participants experienced two simulations: (1) the Flightcrew Procedures Experimental Training (FlightPET) simulation (preflight inspection) developed by UCF as a research tool under part of this project, and (2) the Procedures Trainer developed by Reaction Simulation; using PC-monitor-/tablet-based or VR-HMD-based hardware configurations.

The primary purpose of this study was to test the usability and utility of providing additional training and practice opportunities to pilots as they learn procedures associated with the targeted operations (e.g., preflight inspection, preflight setup). One of our key interests was to capture Part 121 air carrier pilots' perceptions and opinions of different training approaches, which may vary with respect to the modality, interactivity, and level of immersion of that training. Specifically, we were interested in how differences in the level of immersion (low vs. high) and level of functional interactivity (low vs. high) may impact these perceptions and opinions for training for substantively different procedures relevant to air carrier operations.

Based on our prior review of the literature, we were also interested in the variance introduced by various individual characteristics. These individual characteristics could include general demographics (e.g., age range, biological sex, handedness, visual/auditory depreciation/impairment), aviation experience (e.g., flight experience, Part 121 flight experience, B-737 flight experience), and gaming behaviors (e.g., frequency of use, familiarity with controls, preferred genres). In the study reported here, we investigated participants' perceptions of these training approaches by examining pre-post differences in outcomes such as self-efficacy and perceptions of the instrumentality of various training technologies. Finally, we investigated participants' perceptions of and reactions to these training approaches with respect to their simulation sickness, perceived recall, subjective cognitive load, usability, familiarization, and various affective and motivational factors.

4.1. Hypotheses

- *Hypothesis 1.* The VR approach (3D display and peripherals), involving a high level of immersion and high interactivity (spatially and functionally), will be perceived as being more useful / instrumental for procedures involving a higher degree of psychomotor and perceptual skills (e.g., initial preflight setup/flows, unreliable airspeed, go-around), versus the PC training approach (2D display and interface).
- *Hypothesis 2.* There will be no difference between these training approaches, even a preference for the desktop training approach, for procedures more dependent on declarative knowledge (e.g., preflight inspection).

- *Hypothesis 3.* The high immersion conditions will be correlated with more positive perceptions for the initial preflight setup and more negative perceptions for the exterior preflight inspection, compared to the low immersion conditions.

4.2. Method

4.2.1. Participants

Participants in this study were identified and recruited by the air carrier partner. Of the 30 air carrier pilots, 23 were male and 7 were female. The goal of this study was to develop an understanding of how Part 121 air carrier pilots acquire the knowledge and skills to perform procedures relevant to flight operations. In addition, we were interested in capturing pilots' perceptions, reactions, and opinions of different training approaches for flightcrew-relevant tasks. The make-up of the sample along a number of relevant background variables is shown in Figure 11 below. As shown there, the sample's age distribution was varied and encompassed the entire range of 14 CFR 121 pilots: 1 participant was in the age group 21-30 years (3%), 9 participants 31-40 (30%), 8 participants 41-50 (27%), 10 participants 51-60 (33%), and 2 aged 61-65 (7%).

Pilots' total flight time ranged from 4,000 to 23,000 hours (*M* = ~ 9,800 hours; *Mdn* = 8,000 hours; 1st quartile = 5,525 hrs; 3rd quartile = 12,250 hrs). With respect to flight time with a 14 CFR 121 carrier, 2 participants reported having had 1,000 hours or less (7%), while 28 participants reported having had over 1,000 hours (93%). In terms of experience with B-737 aircraft, after which the generic aircraft was modeled, 6 participants reported having no experience (20%), 7 had under 1,000 hours of experience (23%), and 17 participants had over 1,000 hours of experience (57%) with this aircraft. While there is no inherent meaning of this mark with respect to type experience, 1,000 hours of experience is considered a threshold that is (1) often used by the FAA to indicate substantial necessary experience for various qualifications, and (2) equivalent to the maximum number of hours someone could fly in 14 CFR Part 121 operations in one year. Thus, individuals with 1,000 hours of type experience are well qualified and have been with the aircraft for a substantial time and number of operations.

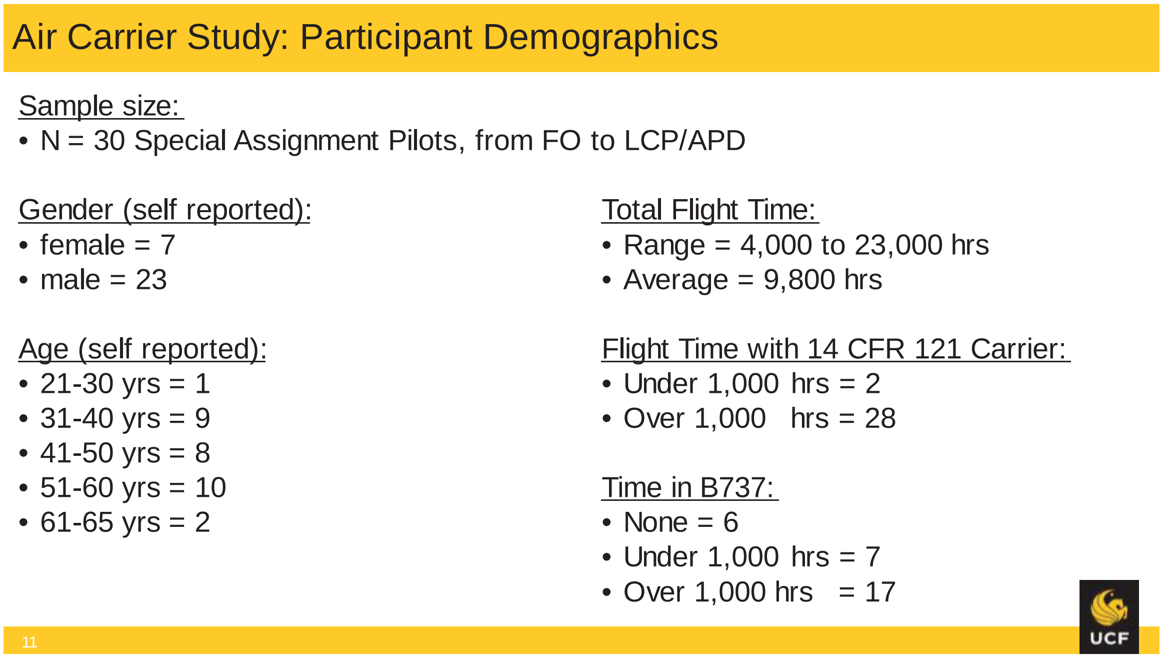


Figure 11. Basic demographic data for the Phase 3 study sample.

4.2.2. Design

This study used a 2 between (Immersion; Lower [2D] vs. Higher [VR]) x 2 between (order: initial preflight setup/flows first; exterior preflight inspection first) x 2 within (Type of Task/Task Spatial Interactivity; Lower [exterior preflight inspection] vs. Higher [initial preflight setup/flows]) mixed-model design, as represented in Table 8 below.

Table 8

Phase 3 Study Design

		Spatial Interactivity/Procedure (Within-Subjects Variable)	
		Lower Interactivity: Exterior Preflight Inspection (FlightPET)	Higher Interactivity: Initial Preflight Setup/Flows (Reaction Simulation)
Immersion (Between Subjects, Independent Variable)	Low (PC 2D)	**PC Simulation with 2D** Initial Preflight Setup/Flows	**PC Simulation with 2D** Initial Preflight Setup/Flows
	High (VR 3D)	**VR Simulation with 3D** Initial Preflight Setup/Flows	**VR Simulation with 3D** Initial Preflight Setup/Flows

Figure 12 below shows the distribution of participants across the between-subjects study variables (i.e., Immersion and Order).

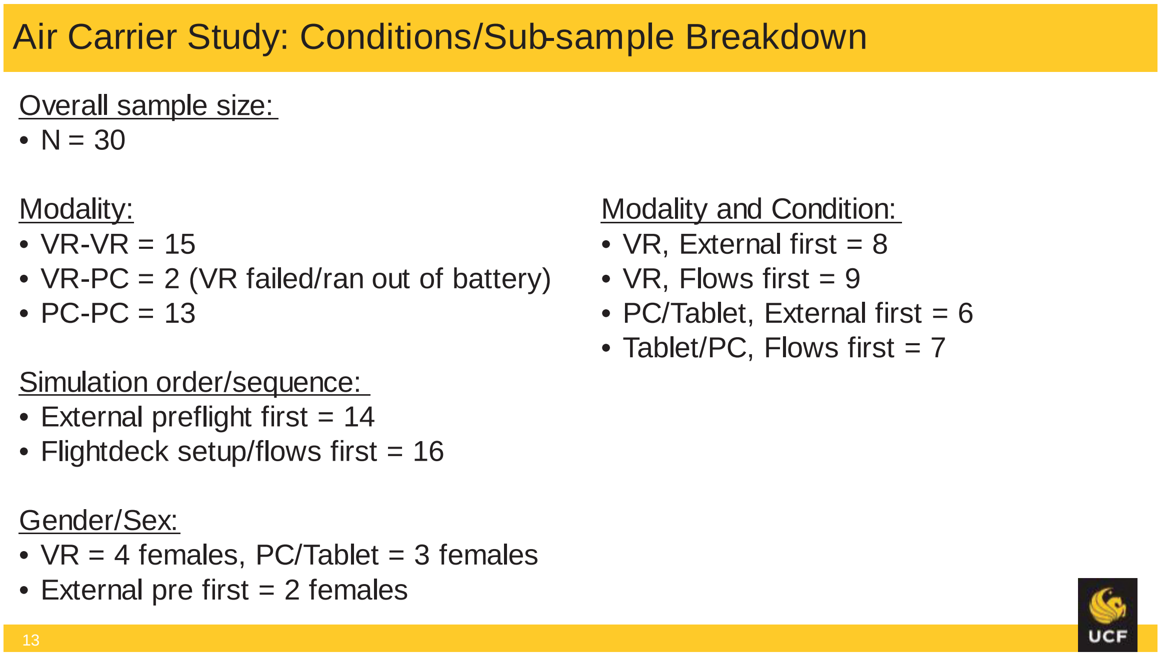


Figure 12. Breakdown of participant assignments to study conditions.

4.2.3. Apparatus

A variety of training approaches, including hardware and simulation software were used for the different study conditions. Examples of the two simulations used are provided in Figure 11 and Figure 12. In each condition, participants received a brief verbal overview of the controls for using that specific simulation, followed by a brief period of system configuration and calibration. Depending on the assigned experimental condition, participants used a standard desktop computer, mouse, and keyboard; a standard computer tablet; or a virtual reality (VR) head-mounted display (HMD) with hand-held controllers, as illustrated in Figure 13.

Figure 13. Instructor settings menu for FlightPET

Figure 14. Reaction Simulation trainer software on a standard tablet

Figure 15. HTC Vive Focus 3 virtual reality (VR) headset and controllers

4.2.4. Materials & Measures

4.2.4.1. *Online Informed Consent Form.* Participants were asked to complete the informed consent form prior to arriving at the air carrier's training center. If participants did not complete the informed consent online, they were consented in-person. This form explained the purpose of the study and any potential risks associated with participating in this study. In addition, participants were told that their involvement was strictly voluntary and that they could decline or withdraw at any point in the study.

4.2.4.2. *Online Pre-Demonstration Survey.* In addition to completing the online informed consent form, participants were asked to complete a series of questions that included participants' age range, biological sex, handedness, and any visual/auditory depreciation or impairment. In addition, we asked questions related to the participants' aviation experience (i.e., total flight experience in thousands of hours; general level/category of Part 121 flight experience, and general level/category of flight experience in a B-737 aircraft) and use of video game systems (i.e., frequency of use, familiarity with controls, preferred genres). Finally, we asked questions about participants' self-efficacy with certain procedures and their perceptions of various training technologies.

4.2.4.3. *Study Conditions.* Participants were randomly assigned to one of four study conditions: (A1) VR Initial Preflight; VR Exterior Inspection, (A2) VR Exterior Inspection; VR initial Preflight (B1) PC Initial Preflight; PC Exterior Inspection, or (B2) PC Exterior Inspection; PC Initial Preflight. Participants assigned to Condition A1 first interacted with the high immersion version of the Initial Preflight procedure, then interacted with the high immersion version of the Exterior Inspection procedure. Participants assigned to Condition A2 first interacted with the high immersion version of the Exterior Inspection procedure, then interacted with the high immersion version of the Initial Preflight procedure. Participants assigned to Condition B1 first interacted with the low immersion version of the Initial Preflight procedure, then interacted with the low immersion version of the Exterior Inspection procedure. Participants assigned to Condition A2 first interacted with the low immersion version of the Exterior Inspection procedure, then interacted with the low immersion version of the Initial Preflight procedure.

4.2.4.4. *In-Person Post-Demonstration and Exit Survey.* After participants engaged in the two training approaches, they completed a series of questions regarding their immediate reactions and perceptions of various training approaches/technologies.

4.2.4.5. *In-Person Discussion & Debrief.* These consisted of semi-structured informal discussion with participants on their perceptions and opinions of the training approaches, followed by debriefing participants on the study purpose and compensation procedures.

4.2.5. Procedure

For this study, participants completed online onboarding activities which included the informed consent form and pre-demonstration survey prior to arriving at the air carriers' training center. Participants that did not previously complete the online onboarding were consented in person. Participants were assigned to one of the four study conditions (2 Immersion [VR vs. PC/tablet] (between) x 2 Orders [external preflight inspection first vs. initial preflight setup/flows first]). After each training demonstration, which, with system familiarization or tutorial, each lasted about 20 minutes, participants were asked to complete post-demonstration surveys related to their immediate reactions and perceptions of the training approaches and technologies presented in this study. They then completed a set of post-demonstration assessments that asked global questions about training technologies. Following this, and towards the end of the study session, participants were asked to participate in semi-structured informal discussion, followed by debriefing participants on the study purpose and compensation procedures. In total, each session took approximately 10 minutes for the pre-demonstration online survey and up to 60 minutes for the in-person portion.

4.3. Results

An initial data analysis of the Phase 3 study focused on three aspects, in accordance with the study's research goals and specific hypotheses: First, we compared perceptions of instrumentality between high- and low-immersion (VR vs PC/tablet), for both tasks (Hypotheses 1 and 2). Second, we compared perceptions of usability/cognitive load across level of immersion, for both tasks (Hypothesis 3). We augmented the numerical results with a review of pilot's qualitative comments related to each simulation and its instantiation. We also, more broadly, looked at perceived instrumentality of different training approaches for each task, both before participation and once they had interacted with the simulation, and we analyzed the simulator sickness (Total SSQ) scores reported by the participants.

4.3.1. Pilot Perceptions of Instrumentality and Usability

Initial analyses of the numerical results from the Phase 3 study showed that most participants reported the VR-based simulation tools as an improvement to their current training methods. Across instantiations, they reported that interactive simulation can be a viable tool that could be used to augment or even replace established training approaches such as paper tiger training, current PowerPoint, and live training. Figure 16 shows a detailed breakdown of ratings of perceived utility and instrumentality that pilots provided towards different learning modalities for initial training of the two task sets.

Air Carrier Study: Instrumentality and Perceived Utility

Preflight Walkaround – all participants:

- *Before/after* experiencing FlightPET:
  - Text (before/after) = 28 / 37
  - E-learn (before/after) = 32 / 49
  - Video (before/after) = 40 / 49
  - VR (before/after) = 69 / 81

Preflight Walkaround – VR Ss only:

- VR (before) = 58
- VR (after) = 85

Preflight Walkaround – PC Ss only:

- CBT (before) = 34
- CBT (after) = 61

Flightdeck Flows – all participants:

- *Before/after* experiencing RS Flow Trainer:
  - E-learn (before/after) = 35 / 46
  - Paper tiger (before/after) = 37 / 35
  - FFS (before/after) = 61 / 63
  - VR (before/after) = 63 / 81

Flightdeck Flows – VR Ss only:

- VR (before) = 56
- VR (after) = 82

Flightdeck Flows – Tablet Ss only:

- CBT (before) = 37
- CBT (after) = 65

Note: All above on scale from 0 -100, in response to "this approach would be useful for training [target topic]"

15 UCF

Figure 16. Summary table of ratings of perceived utility and instrumentality for initial training, both before and after pilots in the Phase 3 study experienced the respective training simulations.

As Figure 16 shows, pilots rated VR highly, both before, but also after they had interacted with the systems. In fact, participants reported higher ratings of the modality with which they interacted after experience with it, as compared to before, whereas the ratings for training approaches they were familiar with (e.g., paper tiger or FFS for initial preflight setup/flows) did not change substantially from pre-participation to after. This suggests that the systems exceeded pilots' expectations.

Test of the a priori hypotheses regarding perceived instrumentality (Hypotheses 1 and 2 above), showed partial support for the hypotheses: The numerical ratings of the usefulness of either form of initial preflight setup/flow trainer did not differ (all $p > .10$) between those who had interacted in VR with it and those who had interacted with the tablet version, thus not supporting Hypothesis 1, which had stated that the VR modality would be rated higher than the tablet version for the initial preflight setup/flow trainer. However, in their open-ended comments, participants very consistently and clearly described VR as a "game changer" for training initial preflight setup/flows and one pilot even summarized their sentiment as a "tool every pilot wants in [their] hotel room, like yesterday". Participants even rated the VR initial preflight setup/flow trainer higher than established methods for training initial preflight setup/flows—including the FFS, as shown in Figure 16. Indeed, several stated explicitly that that they thought VR, more generally, could be used to create low-fidelity simulations of "busy work" tasks, including for the practice of initial preflight setup/flows, in lieu of the more costly FSTD's. Many participants also pointed out that the VR simulation technology would be most appropriate for new hires in initial training, and in other special circumstances for practice, such as after returning from medical/parental leave or other management assignments.

In support of Hypothesis 2, there were no differences in perceived utility of the *exterior preflight inspection trainer* between the two immersion conditions (all $p > .10$ with near-identical mean scores [e.g., 5.06 to 5.15 on a 6-point scale]). That said, those who interacted only with the lower immersion versions (PC-monitor and tablet) rated the potential utility of VR for this task higher after training, than those who actually had experienced the VR.

To test Hypothesis 3, which predicted that the more immersive simulations would be rated as higher for the initial preflight setup/flows and as lower for the exterior preflight inspection, we found partial support: For example, cognitive load self-report data after exposure (CL1) showed a significant interaction of simulation type and immersion, F (1, 26) = 5.994, p = .021, in that there was no significant difference in cognitive load reported for the initial preflight setup/flow trainer by modality (M_{VR} = 6.63 vs. M_{tablet} = 7.20), but a significantly lower cognitive load for the PC implementation of the exterior preflight inspection trainer than for its VR-equivalent (M_{VR} = 6.84 vs. M_{PC} = 5.84).

4.3.2. Simulator Sickness Scores

We also have started to analyze the simulator sickness scores reported by the Phase 3 participants after interacting with the two simulations. Figure 17 shows initial results of the analyses of SSQ scores, here, specifically of the Total SSQ scores after using the initial preflight setup/flow trainer.

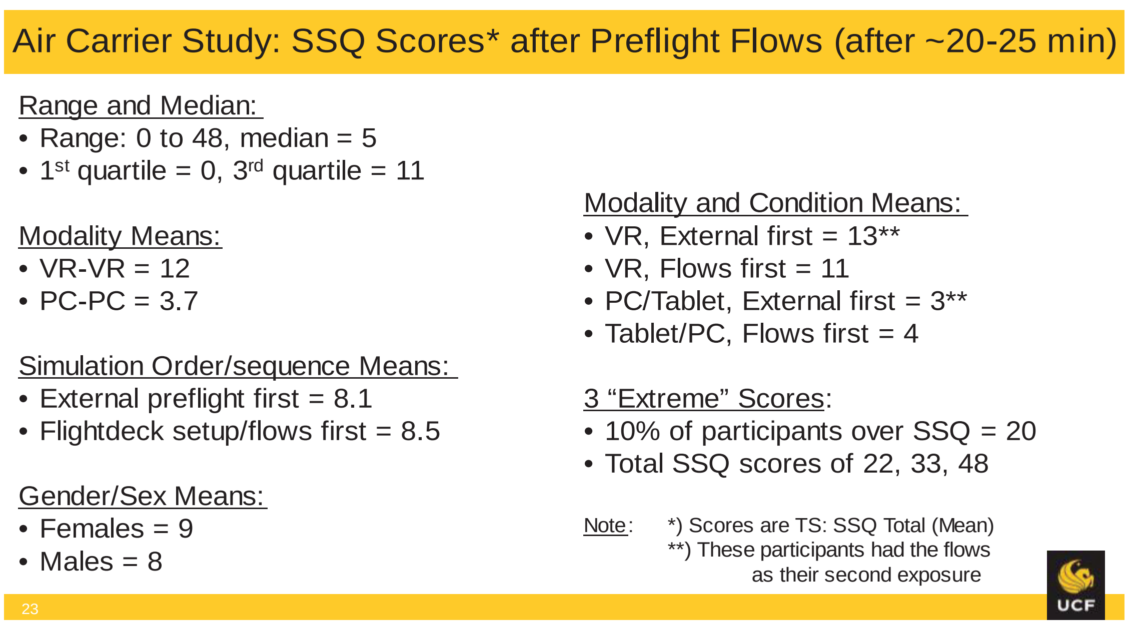


Figure 17. Summary of SSQ scores reported by participants for the initial preflight setup/flow trainer.

A few items deserve mentioning. First, the overall level of simulator sickness pilots reported was comparatively low, after the relatively short exposure period they had, which was no more than 15-20 minutes. That said, second, three of the participants who had experienced the initial preflight setup/flow trainer in VR (i.e., 10% of the total and about 18% of those who had interacted with the initial preflight setup/flow trainer in VR) reported scores above 20; whereas none of the participants did who had used the tablet-version, as illustrated in Figure 18.

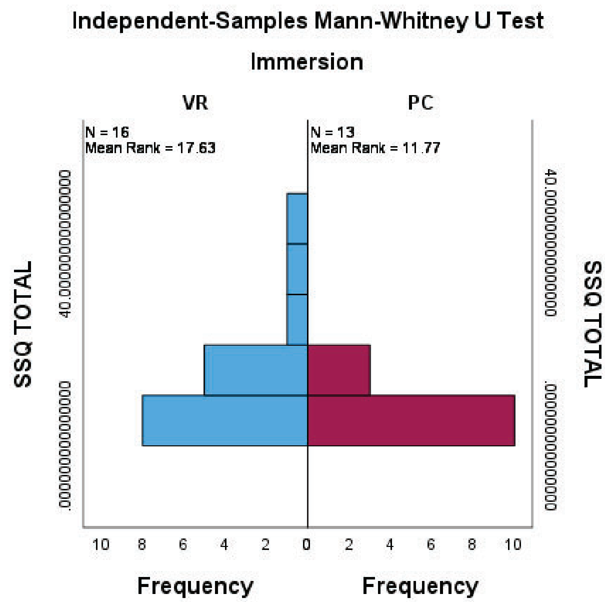


Figure 18. Histogram plot of the frequency distribution of Total SSQ scores in Phase 3 after use of the initial preflight setup/flow trainer by those who had used the VR version (left) and the tablet version (right, labeled "PC").

Third, patterns of results for the SSQ followed expectations: Participants who experienced the initial preflight setup/flow trainer after already interacting with the external preflight inspection trainer, and thus had higher cumulative exposure, scored somewhat higher on cybersickness than those who started with the initial preflight setup/flow trainer. Finally, regardless of order, the VR version elicited higher ratings of simulator sickness than the tablet version of the initial preflight setup/flow trainer, F (1, 27) = 4.428, p = .045 (Table 9); in fact, the reported scores were significantly higher and, in relative terms, the average Total SSQ score was four times as high for the VR group than for the tablet group (M_{VR} = 12.00 vs. M_{tablet} = 3.72), with an effect size of partial Eta^2 = .141 (adjusted .109).

Table 9

ANOVA Table for the Total SSQ Scores after the Initial Preflight Setup/Flows in the Phase 3 Study

Tests of Between-Subjects Effects

Dependent Variable: SSQ TOTAL

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	492.759[a]	1	492.759	4.428	.045	.141
Intercept	1772.225	1	1772.225	15.924	<.001	.371
condab	492.759	1	492.759	4.428	.045	.141
Error	3004.935	27	111.294			
Total	5489.855	29				
Corrected Total	3497.694	28				

a. R Squared = .141 (Adjusted R Squared = .109)

Similar results were found for the preflight inspection trainer. Although none of the differences reached statistical significance with this relatively small sample and the very short exposures, the pattern of SSQ results was the same as for the initial preflight setup/flow trainer; in both cases was the average for the VR condition nearly three times as high as that of the equivalent PC version, the latter of which was constrained downward by the near-zero reports of any sickness symptoms by respondents in the PC condition: $M_{VR} = 11.31$ vs. $M_{PC} = 3.72$ (see Figure 19).

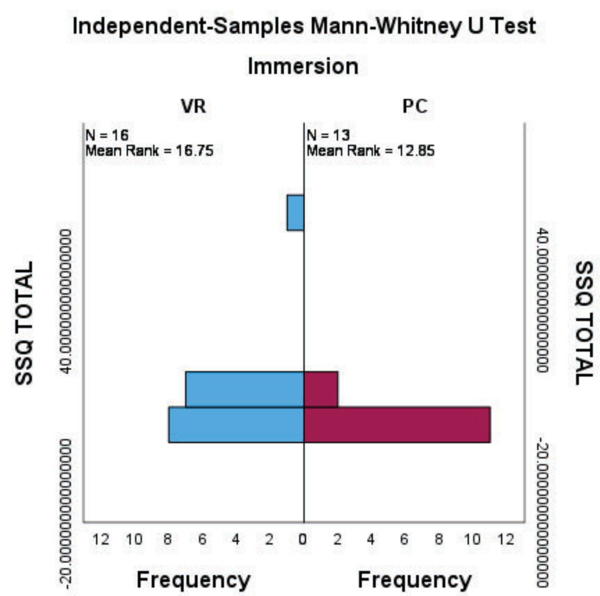


Figure 19. Histogram plot of the distribution of Total SSQ scores in Phase 3 after use of the exterior preflight inspection trainer by users of the VR version (left) and the PC version (right).

5.0. Discussion

5.1. Overall Summary of Results of the Phase 2 and 3 Studies

In Phases 2 and 3 of this project, we conducted two empirical studies that collected data from a total of 132 participants, including 30 active airline pilots employed by a major U.S. air carrier. Specifically, we collected perceptions of usability, assessments of instrumentality and efficacy, measures of learning, and indications of cybersickness from participants who interacted with either immersive VR configurations involving pilot-relevant procedural aviation tasks, or their PC- or tablet-based equivalents. We also collected numerous human-system interaction data from the training system in the Phase 2 study, and engaged in detailed comment capturing in the Phase 3 study; these data continue to be analyzed.

The results of the analyses of both quantitative data and qualitative comments from both studies that both the highly immersive VR- and the less immersive PC-/tablet-based equivalents of the training simulations were generally assessed very positively by both participant groups. Both training simulations were considered improvements over currently available training methods (e.g., paper tiger, training texts for the initial preflight setup/flows; videos or PowerPoint training for the exterior preflight inspection) by the Phase 3 pilot participants, but there were notable differences between implementations in usability, perceived instrumentality, and the level of cybersickness experienced by our study participants between implementations, depending on the task.

The *preflight exterior inspection* simulation, regardless of implementation, was associated with substantial knowledge gains in the Phase 2 study, when compared to the pre-study level of knowledge among the participants. Further, there were no significant differences in either gains or absolute post-test knowledge measures between the VR and the PC implementations, as we had hypothesized; in fact, on average, the PC version was associated with higher, but not statistically significantly greater, knowledge gains. However, contrary to our expectations, the greater interactivity provided by the active learning strategies was not associated with uniformly better learning outcomes: Both the active PC and the passive VR version showed similar and the overall greatest learning. Further analysis is required to determine whether there is an explanation, such as usability or workload, that may explain this. We did observe small differences in usability—the VR solution was rated lower, on average, albeit not statistically significant. We also observed very large differences in cybersickness—the VR solution led to, on average, highly significant and substantially greater levels of cybersickness (doubling SSQ scores from below 20 to near 40, at a medium statistical effect size).

A similar pattern was also found among the air carrier pilots, who were very positive towards both implementations of the preflight exterior inspection training software (i.e., FlightPET), but also indicated that they thought the VR implementation was more cumbersome, probably not needed, and who reported nearly four-times the levels of cybersickness (albeit at very low absolute levels [<20]) after experiencing the VR-implementation for only 10-15 minutes, when compared with the PC-implementation of FlightPET. Post-experience, the Phase 3 pilots also rated the desktop implementation of the preflight exterior inspection software both higher when compared to before experiencing it, and significantly higher than other currently used alternatives, such as videos, PowerPoint presentations, etc. Together, this indicates that, as we predicted, for a procedural task that had a perceptual but limited motor component, that is, the preflight exterior inspection, the use of immersive VR was not justified in terms of acceptance, perceived instrumentality, and effectiveness. Indeed, using immersive VR for this task likely would be contraindicated due to the higher levels of frustration and cybersickness trainees may experience, as compared to the highly rated PC-desktop equivalent.

In contrast, for the *initial preflight setup/flows* training software, which targeted a task with a very active and spatial motor component, participants in the Phase 3 air carrier study consistently expressed their belief that the immersive VR-implementation was "value added", not only over the tablet-based version, but even when compared to the FFS or any other form of training device, from FTDs to the paper-tiger. Despite the fact that those using the VR implementation, on average, also rated their cybersickness levels about 4 times the level of those who interacted with the tablet version (again albeit at very low absolute levels [< 20]) and despite their concerns about having to learn the VR technology, the ability to actively interact with the flight deck, reach for and operate switches, knobs, and buttons, was consistently considered a major improvement by the participants for this highly spatial and bodily active task. That said, it must be mentioned that the tablet version was still very highly rated, based on its interactivity and the opportunity to practice the initial preflight setup/flows when using a VR-implementation would not be feasible. In fact, pilots rated the CBT (tablet) version of the training, after experiencing it, as being as useful for practice as the FFS, especially for initial training and quick review.

5.2. Conclusions: Promises and Pitfalls of XR Solutions for Aviation Training

Together, the extant literature, which we reviewed in Phase 1 of this project, other research since, and the results of our current Phase 2 and Phase 3 empirical studies, suggest that interactive extended reality (XR) simulations, and in particular those which use Virtual Reality (VR) or Mixed Reality (MR) as the display-, interaction-, and delivery-platforms, have the potential to provide training for applied aviation tasks that can be as effective, or even better, than traditional training modalities, while providing substantial potential advantages in terms of training logistics, cost, and customization to the trainee. The possibility of creating immersive, realistic, and interactive representations of job-relevant spatial/3D environments anywhere-anytime, while alleviating the dependence on large, fix-based infrastructure and potentially reducing the scope of dedicated equipment required. Infrastructure, and equipment, which are often constrained in terms of initial cost, time, and space, and of limited scalability. This is an appealing value-proposition for trainers and trainees alike. It is no wonder that the idea of being able to create an airliner's flight deck, its passenger cabin, an air carrier's maintenance hangar, or an airport's tarmac, pretty much anywhere where there is a chair and a 3 cubic meters volume of space, excites the aviation industry, as it promises possibilities for training that is unshackled from limited and expensive training centers or from the real environment, with all its limitations and costs..

Finding 1: XR-environments have the potential to provide equal or better training for applied aviation tasks than current training conducted in other environments, at potentially lower cost, higher scalability, and greater access and customization to trainees.

To realize this potential, however, those embarking on conceptualizing, building, acquiring, using, or assessing XR-solutions for training applied aviation tasks must be aware that a number of factors affect the solutions' usability, efficacy, and effectiveness. While it is true that well-designed XR training that is employed thoughtfully and assessed carefully has the potential of equal or better effectiveness, at larger scales, and with lower acquisition and operating costs per trainee, poorly designed or poorly applied XR training will likely be less effective, possibly more costly, and indeed can potentially be averse to trainees.

For example, an XR-solution that cannot keep up with changes in Standard Operating Procedures that an organization may implement, will quickly be outdated and may become disused. Similarly, an XR-solution that displays an interaction on the flight deck incorrectly, may provide negative training. And if trainees are asked to spend extended periods in an XR environment, without appropriate breaks or careful monitoring, this may result in substantial instances of cyber-sickness, discontinuance of training, and high cost for alternative training and qualification methods for the affected trainees.

Finding 2: An XR-solution that is poorly designed, employed, or assessed, may not reach its potential; indeed, it may be more disadvantageous than the traditional training it is intended to supplement or replace.

Further, in the same way that even the actual aircraft or the highest-fidelity and most expensive Full Flight Simulator (FFS) are not "training," *per se*, neither is an XR environment, if it is not well integrated into a training program, or if it is used without consideration for learning objectives, or if it does not consider the physical limitations of the trainees. All too often, the focus in XR training solutions seems to be mostly on describing its immersion and realism, that is, the replication of the job environment with an unwarranted emphasis on visual representation. However, it is well known that, depending on the training objectives for an applied aviation task, even a table-top exercise, held in a classroom, with nothing more than a set of charts and a few flight documents, can be a better and more effective training environment than the actual aircraft. In those cases, it is the way instructional objectives are translated into training opportunities that makes even an "unrealistic" environment quite effective: Effectiveness emerges, for example, through the way in which instructors facilitate guided practice, provide structured scaffolding, and engage in assessment of learning objectives.

Finding 3: XR and XR-environments are not training, per se. XR environments are a medium for conveying information to pilots. They become training by being carefully integrated into a systematically developed, used, and assessed training eco-system. What information is conveyed, and the setting in which it is conveyed, are equally important considerations..

5.3. Turning XR Solutions into Effective Training

Organizing Framework. We suggest that a useful three-dimensional framework to conceptualize, classify, and compare XR solutions and their integration into the training eco-system considers *immersion* and *interactivity*, as originally proposed for VR by, among others, Slater and Wilbur (1997), and which adds *realism* as a third dimension (see Section 2.5. above; Figure 1). In this context, immersion refers to the degree to which the XR solution provides a sense of vivid presence in the target environment; interactivity to the way the experience is changed from passive to active; and realism describes how the XR experience approximates the existing reality of the target environment.

Finding 4: A useful way to conceptualize, classify, and compare XR solutions is to assess and quantify their immersion, interactivity, and realism.

As we pointed out further above, the three dimensions of XR are largely independent from one another: A VR game may be highly immersive because it creates a vivid sense of presence and highly interactive because it provides many opportunities for interactions with the environment, but it may ignore natural physics, and thus be completely lacking in realism. Conversely, a telepresence VR experience may provide full immersion and stunning realism in 8K UHD, but it may lack any opportunity for interactions. And a systems trainer implemented on a tablet display may fall short on immersion, but it may be highly interactive and be realistic in its depiction of displays.

Finding 5: Immersion, interactivity, and realism are independent/orthogonal dimensions of XR solutions.

Advantages of immersion in XR. Proponents of highly immersive XR displays often advance the notion that immersive XR allows trainees to experience true presence in the environment that is targeted for training. There are two main advantages of high degrees of immersion in XR. The first is that the spatial relationships of the modeled environment are similar to those in the targeted environment: For example, things "behind" the trainee in the targeted environment are also behind the trainee in the immersive environment—as opposed to being to the left or right, or still in front, as when a limited number of monitors is used on desktop applications. Seeing what's behind requires one to turn around or look over one's shoulder, unlike in a desktop simulation, thus also increasing realism. Thus, the spatial and metric relationships of elements in the targeted training environment map closely, perhaps even identically, to those experienced in the target environment. Reaching up to a switch that is located on an overhead panel, allows trainees to experience and practice procedures in ways that are richer, because they involve embodied actions, and allow them to internalize not only the sequence of actions, but also the related gross- and, if modeled, fine-motor activities and patterns. This allows for embodied learning in immersive VR (see also McGowin, Fiore, & Oden, 2022), which also strengthens the connections between different learning elements. Participants in both our Phase 2 and Phase 3 studies consistently reported that, for procedural tasks that involve substantial motor patterns, such as the initial preflight setup/flows, this was significant value added – in fact, several referred to it as "game changing."

Finding 6: High degrees of immersion, when paired with realistic spatial relationships in XR solutions, allow trainees to experience and internalize not only the sensory experiences as in the target environment, but also to practice the necessary motor activities for looking around and for reaching and manipulating objects, thereby building motor patterns, and to engage in embodied learning.

Second, the XR training environments that are currently on the market or in development, including the two we used as part of the Phase 2 and 3 studies, are typically fully immersive and model all 360 degrees surrounding the trainee. Thus, unlike in a full flight simulator, when trainees turn around in an XR flight deck simulation, they see the jump seat, the flight deck door, and the bulkhead, rather than the simulator's instructor or observer stations and a bunch of computer closets. Further, as one moves down in the fidelity spectrum from FFSs to FTDs (and FNPTs [EASA]), to AATDs and BATDs, the "share" of space taken up by non-relevant features increases further. Such non-relevant features could include, for example, empty panels, training room walls, flooring, or workstations. High immersion not only accurately represents the actual relationships of spatial elements in the target environment, it also increases the degree to which presence is felt, by shutting out the physical reality of the training location that is unrelated to the training environment. This is potentially important from a training effectiveness perspective whenever suspension of disbelief is critical. Admittedly, suspension of disbelief and presence may not be important for all training evolutions, but they can be important if there is a psychological, specifically affective/emotional, component required for training, such as to provide credible stimuli for training of startle or surprise, or to invoke a sense of danger for decision making training under stress.

Finding 7: By replacing the stimuli of the physical environment where training takes place with stimuli generated from a virtual environment, highly immersive XR technologies may be critical, and indeed superior to other training systems, when suspension of disbelief is important to achieve training objectives, especially in situations where affective components, such as danger, stress, or surprise need to be generated.

The opportunity to bring, through high degrees of immersion and realism in XR, relevant training environments to "*anywhere-anytime*", without the need for physical training facilities and specialized stationary equipment, is of course one of the main value propositions of XR. High immersion not only increases situational interest and improves presence (Peterson et al., 2023; Sonnenfeld, Nguyen, Gomez, Jentsch, & Fiore, 2023), which improve, by themselves, learning, the ability to create training anywhere-anytime also opens up opportunities for conduct training where it previously did not exist.

While our Phase 3 participants, all of whom were active pilots, were split with respect to whether they wanted to have access to XR training only at the training center or also at home/their crew hotel/domicile (more on that below), they did mention and appreciate the opportunity that XR provides to create valid training in any room, on demand, as one of the major draws of the technology. They particularly believed in the value of additional practice opportunities anywhere-anytime for initial, transition, and upgrade training—although they were somewhat less optimistic for recurrent training. Multiple participants also mentioned the opportunity to practice anywhere-anytime in the context of pilots who, for medical, professional, or other reasons, had been away from line flying for a while, and would be interested in practicing tasks before coming to the training center, going into the simulator, or re-joining line operations.

Finding 8: Highly immersive XR solutions allow relevant training environments to be experienced anywhere-anytime, while alleviating the dependence on large, fix-based infrastructure and potentially reducing the scope of dedicated equipment required. Especially compared to non-immersive training systems (e.g., desktop or tablet-based training), immersive XR can also reduce distractions typically present if training is held in multi-use or residential settings.

Another related use of immersive XR solutions for training, albeit currently not yet implemented to its full potential in extant aviation VR solutions is the opportunity for trainees, who are not physically co-located, to experience the same shared virtual training space through telepresence. That is, unlike in a traditional training device such as an FTD, the trainees, and potentially the trainer(s) as well, could meet and interact within the same virtual environment while ostensibly being physically in widely dispersed locations – limited only by technical considerations such as network bandwidth or transmission lags, financial limitations such as access to high-end XR displays, or practical concerns such as diurnal patterns across different time zones.

Remote synchronous collaboration is of course not limited to XR environments, and it is already prevalent in distributed multiplayer gaming, virtual conferencing, and shared remote use of desktop productivity applications, but it has particular appeal for crew training in aviation, which heretofore was limited to physically co-located training in a flight simulation training device at the location where that device was installed. In addition to the potential economic and environmental savings that distributed training would provide, it may also reduce the need for current adaptations to real-life scheduling problems, such as the use of "seat filler" pilots in training when an originally scheduled trainee is unavailable to complete a crew.

Further, even if the possibility of distributed crew training and practice for credit is not executed, giving trainees the opportunity to informally "meet" in a virtual flight deck to practice, while being physically distributed, may provide important opportunities for peer-mentoring and informal professional development. This was something that several of our Phase 3 participants mentioned, again usually in the context of new hire pilots or related to pilots who, when re-joining the line, would be interested in practicing tasks and receiving peer mentoring.

Finding 9: Highly immersive XR applications present new opportunities for joint training of physically distributed trainees and trainers. Unlike in traditional FSTDs, which are limited by the location of the physical device, participants could be accessing the training environment and interacting with it synchronously, but from very different physical locations.

Disadvantages of XR immersion. Although high degrees of immersion can be desirable for effective training, there are also some disadvantages to consider. First, the degree to which highly immersive systems "shut out" the actual physical environment surrounding the trainee, they also take the trainees' attention and focus away from the physical reality in which they are located. This is desirable for creating presence, but it can lead to potentially adverse situations, such as when trainees are colliding with or hitting objects or people in the physical environment while interacting in the virtual world, when trainees are may be prone to tripping or falling in the physical world, and/or when they become unaware and/or unable to react to external events like fire, flooding, or intruders that present themselves in the physical environment.

Finding 10: To the degree that highly immersive XR solutions "transport" trainees away from their physical reality, necessary interactions with, or in, the physical reality where training takes place will become more difficult.

However, even if these events appear drastic, far-fetched, or unlikely, less severe negative outcomes are possible, for example in terms of training or classroom management. An instructor overseeing a group of 10 trainees, each using a XR headset, while they are co-located in a classroom, may have their hands full in getting individual trainees' attention, supporting them, and troubleshooting issues, and there may be a substantial additional load on trainers in those settings, compared to traditional classroom settings. This was also a concern that our experimenters, as well as our Phase 2 and 3 participants, consistently mentioned: Frequent donning and removal of XR headsets, which may become necessary to interact in the physical world, can become a barrier to training and learning, and related frustrations among trainers and trainees alike, can negatively affect the learning environment.

Finding 11: Highly immersive XR solutions may make interactions between and among co-located trainers and/or trainees, more difficult than less immersive alternatives. That is, it is more difficult to interact with other individuals in the surrounding physical environment when the senses are immersed by virtual stimuli.

In this context, it is also important to acknowledge that, while highly immersive systems may allow for spatially distributed trainers and trainees to interact remotely in a shared virtual space, the representation of the trainees in the virtual world, typically through avatars, may lack critical cues, including facial expressions, gestures, and even the physical similarity of trainees with their avatars. At this point, while joint crew training in a shared virtual environment may be suitable for practicing procedures, executing checklists, etc., it is unlikely that the richness of real crew interactions in face-to-face joint training, such as in an FSTD, can be displayed with sufficient accuracy in an XR solution, unless the trainees are co-located, and blended MR using sensor-fusion is employed. The description for Finding 18 below, provides an example of MR relevant to this context.

Finding 12: While highly immersive environments have the potential to be used for crew-level training, the lack of accurate avatars with realistic representation of appearance, nonverbals, and facial/gaze cues, currently limit the utility of VR for tasks involving crew coordination and communication.

Finally, highly immersive XR solutions can only be effective to the degree that the mapping of avatar representations in the virtual world to the physical location and pose of the trainee is correct. Current VR systems continue to experience drift and calibration issues. Further, the representation of "self" in the virtual world can be affected by such technical issues, thereby quickly reducing the sense of presence.

Ideally, trainees should not experience mismatches between where they and parts of their body are in physical space, and what is represented in the virtual environment, as this (a) leads to frustration,

(b) increases the likelihood of inadvertent actions, and (c) potentially induces cybersickness. In this respect, slightly less immersive environments, which blend elements of the physical and virtual worlds, such as in MR, using see-through systems or sensor fusion, may be preferable over full VR that only includes virtual representations.

Finding 13: High degrees of immersion require consistent and reliable mapping of the relationship between the physical location and pose of the user and its representation in the virtual world. Drift and a frequent need to recalibrate, can reduce presence, increase frustration, and induce or exacerbate cybersickness.

Advantages of interactivity in XR solutions. A consistent observation from our empirical studies concerned the importance and desirability of interactivity of training. Trainees, regardless of whether they were professional air carrier pilots or college student participants, and regardless of whether they interacted with the VR environments in our studies or their equivalent PC/tablet implementations, consistently emphasized the importance of being able to interact with the training and actively explore the learning environment. They wanted to "do things" and see "what happens in response." This maps against two, related, but somewhat independent aspects of interactivity, i.e., the "what" of which actions and consequences are modeled, on one hand, and the "how" of which modalities are used to perform the actions.

In a Flight Management System Trainer (FMST), for example, the types of interactivity with a Control/Display Unit (CDU) may be limited by the number of input or line select keys that can be actuated, or by limiting the functionalities that the FMC models (say route inputs only), or by restricting the navigation data base (to certain airports). This determines the "what" of interactivity (and is often closely related to a simulation's functional fidelity). The second aspect is the "how" of interactivity, and here a PC-based FMST may restrict interactions to input via an external mouse (this is often related to a simulation's physical fidelity).

While there are psychomotor tasks that require both interactivity and high physical fidelity of the interaction (see also Finding 18 below), participants in our studies consistently emphasized the importance of the former (the "what") in interactivity over the latter (the "how").

Finding 14: There are two aspects of interactivity relevant for XR systems: the degree to which interactions with the training environment are possible and consequences are displayed ("what"), and the way in which these interactions are implemented ("how"). The "what" is at least as important as the "how".

The desire to actively engage with a training environment and to observe the consequences or outcomes of actions is in line not only broadly with the learning and training literatures (cf. Kaplan et al., 2020; McGowin et al., 2022), which emphasize the importance and benefit of active and embodied learning, it also specifically matches recent results from other empirical research, for example, Peterson et al. (2022), who investigated the combined effects of immersion and interactivity in VR. Peterson et al. concluded that:

"…both immersion and interactivity positively influenced the levels of physical presence experienced by learners, while interaction effects showed that interactivity is more important for

the experience of agency and embodied learning under conditions of low immersion. Moreover, significant main effects suggested that high interactivity reduces extraneous cognitive load from the environment, and that high immersion leads to larger situational interest" (Peterson et al., 2022, 104429).

Finding 15: High interactivity is not only a characteristic highly desired by users for a training solution, but it also improves learning, especially under conditions of low immersion.

The finding by Peterson et al. above that interactivity is particularly helpful when immersion is low is also something that our Phase 3 participants intuited in their assessments and reflections on the PC and tablet versions of the two training systems they interacted with. While the VR solutions were generally higher rated, especially in terms of perceived utility, than their PC/tablet equivalents, the PC/tablet versions were consistently much higher rated than current approaches to training the two sets of procedures that did not include interactivity, such as the paper tiger for the initial preflight setup/flows, or instructional videos or PowerPoint slides for the external preflight inspection. Almost uniformly, participants commented positively on the interaction opportunities that the PC/tablet versions provided them. They also stated that, if only one or the other was available, they would prefer greater interactivity over greater immersion.

Finding 16: When trade-offs must be made between immersion or interactivity, increasing interactivity may be able to "make up" for loss in immersion, especially for tasks where being able to observe the consequences of actions is of utmost importance for training. For tasks such as initial preflight setup/flows and external preflight inspection, participants considered interactivity as more important than immersion.

Selecting interactivity for XR solutions. Current XR solutions differ widely in the degree to which they provide a passive vs. active experience, in the same way that other media and training systems differ in interactivity. As such, selecting the degree of interactivity and matching it to the objectives for the XR solution should, in principle, follow the process of any systematic training needs analysis and design, for example, as described in AC 120-54A CHG 1 (FAA, 2017); for some training objectives, passive experiences such as instructor demonstrations or instructional videos, may be suitable, whereas for others, active experiences, such as interactions with flight deck systems using an FTD, may be appropriate and/or required. As these principles are described elsewhere and not unique to XR solutions, we will not describe them here in detail.

Finding 17: The interactivity of aviation training XR solutions should be determined based on the training objectives, in the same way that levels of interactivity are already selected/specified for other training systems and devices. Extensive guidance material on this topic already exists, and it can be applied to XR systems.

Similarly, what aspects of interactivity should be represented, and how interactivity is implemented in XR systems is not inherently or fundamentally different than for other training systems. For example, the aforementioned FMST (see Finding 14) may, depending on implementation, differ in the interaction modalities that are represented, such as for entering data into the CDU: A PC-based FMST may use mouse-keyboard entry, whereas a BATD or AATD FMST may employ touch screens, while the FMST portion of a higher-level FTDs may use physical replicas of the actual CDU, as found in the aircraft.

Each of these interaction modalities has an equivalent in a typical XR flight deck trainer: In pure VR, trainees may make inputs by pointing at and "lazing" keys from a distance (equivalent to mouse inputs), or by reaching and "pushing" keys (equivalent to actuating a touch screen), whereas in MR or AR, they may interact with physical replicas of the controls (as in a FFS). MR systems such as Collins Aerospace's Coalescence™ system (Collins, 2023), for example, fuse the images of targeted elements of the physical world, obtained from see-through cameras on the HMD, with representations of an external virtual world, using computerized sectoring and/or green-screen technology. This allows users to engage in natural manual interactions with physical objects, such as aircraft controls, within a limited segment of the virtual world, without the need of extensive physical simulations of the remaining environment, which is instead represented virtually within the HMD only.

Finding 18: The selection of how interactions are implemented in XR systems parallels that of selecting interaction modalities for other training devices. Within the XR spectrum, equivalents exist to all common human-system interaction modalities, and similar processes should be followed in deciding what is acceptable and/or required.

Usability aspects of interactivity in VR. In immersive VR, current technologies for the tracking of user orientation, pose, and inputs—especially that associated with the location and orientation of gaze and hands— typically require users to interact with the virtual world using wand-like input controllers in the physical world. At this point, wand-less tracking and pose estimation of hands is limited to a few devices; the vast majority still use wands.

Holding a wand and interacting through traditional input devices on the wand, such as triggers, buttons, track pads, or joysticks, requires the translation of the physical manual interactions in the real-world flight deck into unique input actions via the wand that represent them. Learning the input actions increases cognitive load, requires practice, may lead to trainee frustration, and, depending on the required psychomotor skills, may not be acceptable for training, for example, for training fine motor skills or those with long acquisition curves. Further, there is no standardization across devices of how this is done, or which inputs on the wands do what, potentially further exacerbating the problem (see also the concept of "VR-overhead" further below, in Findings 25 through 28)

Finding 19: In VR systems that use wands for tracking the position and pose of hands and that employ control components on the wands as input devices for interactions, natural manual actions in the real work environment need to be translated into VR-specific actions using the wands. This may necessitate initial practice that is VR-implementation specific, requires time, and increases trainees' cognitive load.

In our studies, the most difficult and frustrating action in VR for most participants was actuating controls by pushing them with a (virtual) finger, such as pressing a virtual CDU keyboard key with one's index finger. In the VR device we used, this required holding the wand in a particular orientation, one that the participants thought was somewhat unnatural, with the physical thumb—rather than index finger—angled towards the targeted virtual push button or key. Then, participants had to move their hand in the physical world towards the representation of the button in virtual space, before they could "press" it. This proved surprisingly difficult for participants: They would miss small buttons entirely, they would not push far enough, or too far, or too quickly, or too slowly. The lack of feedback and success led to frustration and frequent delays. While these were not unlike the problems and frustrations users experience with touch-sensitive displays, the problems participants had with this relatively simple action substantially distracted from learning the locations and sequence of key presses and the overall perception of the training system.

Conversely, an action that we thought was at least as complex as pressing a key, such as grabbing and then moving switches or knobs, proved surprisingly straightforward for participants. In order to actuate a virtual switch or button, participants had to move their virtual hand until it "touched" the targeted one in the virtual world. They then used the wand's trigger to execute a pinching movement of virtual thumb and index fingers to grab it. Once they had grabbed the switch or knob, they then quickly learned how to move the wand in approximately the same way as one would move the physical hand (slide or twist) to actuate the switch or knob. In fact, unlike in the physical world, as long as they held down the trigger, the virtual hand remained glued to the switch or knob, and therefore, no one's hand "slipped" off the switch or button. To release the pinch, participants had to release the trigger, which most mastered very quickly.

Finding 20: How wand-based interactions with virtual objects are implemented is currently not standardized. Mapping of physical actions to VR inputs should not only be considered in terms of replicating/approximating the physical actions of the task environment, but also in terms of the resulting usability. Testing with users may be required, as small differences in implementation may be associated with large differences in usability.

In today's typical VR implementations, i.e., those not using haptic gloves or those without direct input into realistic controls (as in the aforementioned Coalescence™ MR system), realistic haptic feedback is typically missing. In lieu of the haptic feedback that the physical world provides, developers should strive to replicate other forms of actual feedback in the target environment, such as a visual indication that a switch has changed positions, or the illumination or extinguishing of an indicator light when a switch is moved, or an audible "click" as a switch changes position.

Further, the nature of the virtual environment affords opportunities for additional feedback not present in the physical world: As the virtual hand reaches a control, the button, knob, handle could be highlighted, to indicate that it is now reachable or grabbed, and, as it is moved, a haptic signal (such as a brief vibration or a click) could be transmitted through the wand to indicate that actuation took place. Similarly, other auditory or haptic cues should be considered to indicate that a particular action was successfully completed. The vast majority of our Phase 3 participants commented positively on having such additional visual, auditory, or haptic feedback available, and a number of developers of latest-generation XR devices have put emphasis on creating haptic cues for wand-based interactions (e.g., Meta Quest 3)

Finding 21: To compensate for the lack of haptic feedback for VR-based manual interactions with objects in the virtual world, developers should strive to provide other forms of feedback. This includes replicating realistic feedback in other modalities (visual, auditory), and/or adding feedback to the virtual world that is not present in the real world but makes it easier for users to determine that their actions were successful. Extant literature on effective feedback techniques should inform the design of instructional feedback in XR simulations.

Realism and fidelity in XR solutions. Although the discussions so far have focused specifically on immersion and interactivity of XR solutions, the reader will have noticed that the realism of the training environment had to be considered at times – because the purpose of XR training systems in aviation is to prepare individuals for real tasks, a minimum level of realism is inherently required. That said, there are situations where a lower level of realism in certain aspects of the task is acceptable. Indeed, sometimes lack of realism is actually desirable in training: Examples include training of certain transportable skills in analog environments (we mentioned CRM training using a table-top exercise before), visualization of certain processes or functions (such as systems training that uses simplified diagrams), part-task training (that simplifies the complexity of the complete task environment), above-real-time-training or artificially slowed down training, computer-guided practice that adds cues and clues to the task environment, etc. In all of the aforementioned cases, realism of specific aspects of the task environment is intentionally reduced, often because an environment that is exactly like the task environment is too complicated, too difficult, too expensive to replicate, etc.

Finding 22: The required level and elements of realism in XR training solutions should be driven by the instructional objectives. There are valid instructional reasons for intentionally reducing realism in training, to make training available, acceptable, effective, and/or efficient.

Given that matching the realism of a training intervention, device, or content to the training objectives is not unique to XR, we will not review processes and criteria for that here. However, instead, we would like to reiterate (a) that even the 100% real target environment is not training, *per se* (see also Finding 3 above), and (b) that we believe one of the exciting possibilities of XR is their potential to become training solutions that go "*beyond reality*", thereby greatly increasing their utility in training.

For example, consider how the FlightPET simulation of the external preflight inspection created for this research goes beyond reality, to improve training: FlightPET includes visualizations such as highlighting, magnifying, virtual disassembles, dynamic comparisons, and functional animations for selected parts and systems of the aircraft that need to be visually inspected as part of the preflight. Trainees can experience not only what a worn tire looks like, *in situ*, as compared to a new one, they can see, through a disassembled view, why a particular pin left in the nose gear assembly would restrict its movement, or how the airflow into the engine is restricted by cowling icing, via an animation – and all on demand and at a moment's notice. These are opportunities that even the use of an actual aircraft would rarely, if ever, afford, and such instructional elements were consistently rated highly and positively mentioned by our Phase 2 and Phase 3 participants.

Similarly, our Phase 3 participants frequently spoke about how an initial preflight setup/flows trainer could include animations of aircraft system functions, such as showing, using highlighting and system diagram overlays, which systems are powered by different buses and actuated by different switches, or how a fuel imbalance results if the cross-feed valves are not in the correct position, etc. They were excited to imagine the combination of immersion, interactivity, and realism, anywhere-anytime, without the need of expensive system hardware, with such novel instructional contents.

Finding 23: One of the exciting possibilities of XR solutions is to combine highly immersive and interactive environments with instructional interventions that go "*beyond reality*" and allow trainees to experience training environments which not only provide training in relevant settings, but also augment those relevant settings with additional instructional content, such as guided practice and scaffolding, visualization and animation, above- or below-real-time training, and automated feedback.

Ultimately, our Phase 3 participants also acknowledged that aviation is a business and training is costly. Many acknowledged that XR simulations might allow them to learn and practice critical tasks at lower cost outside of otherwise very expensive high-fidelity devices, such as FFSs. More than one of them acknowledged in their comments that "*I don't need to be in a Level D simulator to practice cockpit flows*", as long as a "sufficiently good" alternative was available. In fact, as a group, they actually rated the potential and suitability of the VR-based initial preflight setup/flows trainer higher than that of a FFS, for initial training. Instead, they expressed the hope that FFS time could be used to explore safety-critical behaviors of the aircraft, such as high-altitude maneuverability, upset prevention, edge-of-the-envelope performance, or unusual system failures.

Finding 24: A highly interactive and sufficiently immersive and realistic XR simulation that includes instructional elements may be preferable over training in a higher-fidelity device that does not include those instructional elements. This also can free up high-fidelity devices, such as Full Flight Simulators, for other training and practice opportunities.

5.4. Incorporating XR Solutions into Training

In addition to obtaining critical information on how XR solutions should be specified, designed, and created for maximum effectiveness, we also gained critical insights from this project related to how XR solutions should be incorporated into air carrier training. Considerations included where and when to introduce, what some barriers to adoption and use are, and how training managers and trainers can support the introduction. In the following, we review these considerations and provide related Findings.

"VR-overhead." A main observation from both the scientific and practitioner literatures, from reports in the popular press, and from our own experiences throughout this project and especially in the execution of Phases 2 and 3, has been that there is a considerable amount of what we term here "VR-overhead." That is, to employ XR (currently mostly VR), whether as a developer, deployer, or end user, still requires a lot of specialized know-how, and that know-how is not yet pervasive among the respective constituent and stakeholder groups. Even ostensibly basic tasks, such as how to don and calibrate a headset, how to start a simulation, and how to interact with the simulation hard- and software, require unique knowledge, skill, and experience.

Finding 25: Currently, using an XR solution still requires a lot of specialized know-how, and this know-how is not yet pervasive. This leads to a substantial amount of "VR-overhead" needed for simply using an XR training solution. VR overhead is independent of other, traditional aspects of employing and using training.

The reasons for high VR overhead are manifold. First, at this point, the XR-industry continues to be in an early stage of technology maturity and has neither standardized, nor consolidated. As a result, despite some common approaches to solving similar technical challenges, there is wide variation in the characteristics of the hard- and software in the XR industry. Unlike for desktop, personal, and mobile computers, for video game systems, for mobile phones and tablets, or for wearables, all of which in many ways had comparable early trajectories, the XR landscape continues to be characterized by many entrances and exits, as well as very different, bespoke, and proprietary hard- and software solutions. This has led to exciting new developments and much creativity, but it also makes it difficult to assess compatibility, long-term supportability, and usability of XR solutions. It also hinders any attempt to compare one XR solution to another. For example, not all XR software works with all HMDs, and certain HMDs may not be supported at all, or over time. Not all HMDs have similar or the same compatibility with computing systems. Input devices and button mapping are not standardized. Also, as a result, many XR solutions are unique combinations of computing, display, and interactions hardware and software, and comparing them or generalizing from one to another is difficult, if not impossible.

Finding 26: XR is still in an early stage of technology development, characterized by a lack of standardization and consolidation. As a result, each combination of hard- and software is somewhat bespoke. This hinders assessing and comparing solutions. It also makes it difficult to generalize results and provide minimum standards for XR solutions.

Second, experience with XR is still very limited in the population at large, including among pilots. In addition, even if individuals have experienced an XR solution, the unique nature of every hard- and software combination may not have given them the relevant experience, knowledge, or skill to design, build, acquire, assess, employ, or use another one, even within VR solutions. Having used device A to experience the visuals of an underwater reef may be vastly different to having played a sword-fighting game in device B; and neither may be particularly informative for deciding about, and using, device C with a bespoke training solution for initial preflight setup/flows. Additionally, because each hard- and software combination is different, the potential for negative transfer exists.

Finding 27: Prior XR experience is limited in the population, and the lack of standardization among XR solutions limits the applicability of any prior XR experience towards a new XR solution.

Given the fact that specialized knowledge and experience is required, that standardization does not exist, and that the relevant experience is not yet widespread, the development, acquisition, deployment, and assessment of any XR solution need to account for considerable "VR-overhead": Training on how to use basic elements of the XR system may have to be provided, and technical support needs to be accounted for. At this point, XR solutions, unless they dedicate one unique hard- and software combination to one specific training (and only that training), are not yet "out-of-the-box ready" and definitely not yet "plug and play". Further, every new training solution may again require specialized training and support.

Finding 28: For the foreseeable future, every new XR training solution that is introduced will require dedicated and specialized training and user support to account for VR-overhead.

Instructional support of XR solutions. In addition to providing instructional and technical support to overcome the VR-overhead inherent in using XR technology, an XR training solution also needs to include instructional support and guidance to be effective. Especially if an instructor-less implementation is foreseen, adequate tutorials, help files, lesson plans and paths, reset options, and save options are needed. In this, XR simulations are not inherently different from other computer-based training or simulation systems, but because of the lack of familiarity with XR technology, and because XR devices are not (yet) as suitable for the simultaneous display of multiple information sources (such as via multiple tabs or windows) as traditional computing, that these elements are present and usable takes on additional importance. The pilots in our Phase 3 study, for example, highlighted the need for high-quality tutorials and help features (e.g., wand button mapping), especially for the XR versions, and a consistent desire to be able to leave simulations, save their progress, and then return to that part of the training, rather than having to start anew.

Finding 29: As is the case with other forms of computerized training and instruction, instructional support is critical for the successful use of an XR simulation in training. Elements that should be considered include tutorials, guided practice, help features, and the ability to save progress and/or join specific segments within the training.

Participants in our Phase 3 study, all of whom had extensive experience in receiving aviation training (and several of whom also had extensive experience as trainers), consistently emphasized the importance and benefit of interacting with a human trainer, especially when a training topic or tool is new. Even though both systems they interacted with, regardless of modality, contained computerized guidance or a tutorial, they believed the guidance and feedback that a human trainer can provide would be of utmost importance for realizing the systems' potential. They allowed for self-guided or self-paced practice later, but for initial training, they believed that the availability of human assistance was critical. In fact, several invoked procedures trainers as a useful analog example; while these trainers can be great for skill acquisition and practice, they all felt these trainers were not particularly useful if employed as "sand box" environments only.

Finding 30: New users of XR technologies desire, and benefit from, being guided and aided by a human trainer when using XR training systems, especially at initial use.

Where and when to introduce and use XR. Our Phase 3 participants shared a number of thoughts on the practical implementation of VR technologies in an air carrier environment. As we indicated in prior sections above, they considered whether VR technologies could be an ancillary or replacement for training that did not need to be performed in expensive FFS or FTD simulations, expressed opinions about whether VR technologies should only be used in structured and scheduled training events, as opposed to being available for self-paced practice, and consistently emphasized that VR technologies should not be introduced without sufficient familiarization with the technology and outstanding technical and instructional support. They also expressed fairly consistent opinions related to who would benefit most from VR-based training, specifically, many of the experienced participants commented that VR would be most appropriate for new hires and for initial training.

There was somewhat less consistency among participants' views with respect to whether VR devices could or should be issued to individual pilots: While many participants seemed excited by the idea of being able to bring a VR device home or to a hotel to practice, others feared that it would shift training into their non-work lives, or expressed concern about the level of support that they would receive when accessing VR technologies from locations other than the training center. Finally, some expressed concern about being responsible for maintaining the equipment, updating it, etc. Instead, those participants indicated that they thought the best way to introduce VR would be to initially limit implementation of VR technologies to relatively controlled environments (e.g., training center classrooms, briefing/debriefing rooms, dedicated VR rooms at the training center, Chief Pilot's Office, etc.).

Finding 31: Training organizations interested in implementing VR technologies should carefully consider how and where to introduce the training. A phased approach, with limited small-scale trials, potentially on a volunteer basis, that collects data on usability, availability, and acceptance of the technology, can provide important information to increase acceptance and success.

5.5. Cyber-/VR-Sickness

Description and incidence of Cyber-/VR-sickness. One of the concerns related to the introduction of XR systems into a training program is cyber-/VR-sickness (referred to generically as cybersickness from here on). Similar to simulator sickness, cybersickness is a multi-symptomatic experience of discomfort and physiological symptoms during and after the use of a computer-/display-combination, and, specifically, an XR device. The experience of cybersickness is characterized by symptoms, usually grouped into clusters such as disorientation, oculomotor, and nausea (Kennedy et al., 1993), whose onset can be gradual or sudden, whose severity can be minor or major, and which, depending on severity and circumstances, can persist briefly or potentially for hours after discontinuance.

Some level of cybersickness is a frequent and common occurrence, and both the scientific and the popular literature suggest that, depending on the type or XR experience, its duration, and other factors, it is not unusual for somewhere between 1/3 and 2/3 of all participants to experience at least some discomfort related to cybersickness. Additionally, about 10%-15% of participants may be highly susceptible to cybersickness, and they may experience substantial levels of cybersickness symptoms even in situations where the majority of participants only experience mild or indeed no cybersickness.

Finding 32: Most, perhaps all, people are to some degree susceptible to cybersickness, but about 10% to 15% of the population are particularly susceptible. Training organizations should be cognizant that highly susceptible individuals may be found in any group of XR-users.

Our observations from the Phase 2 and 3 empirical studies confirmed this: Even with the relatively brief exposures in Phase 3, three of the participants (10% of all 30 participants and ~18% of the 17 participants who experienced a VR simulation) reported SSQ scores after the initial preflight setup/flow trainer that were substantially higher than the median score in the combined sample (median = 5) and had, in absolute terms, been considered problematic (i.e., >20) by the developers of the SSQ (Kennedy et al., 1993). Indeed, the highest Total SSQ score was 48, after a comparatively limited exposure and considering that the participants were all professional pilots, who typically show, as a group, lower susceptibility and greater adaptation to motion- and simulator-sickness.

Similarly, in our Phase 2 study, which featured a larger, but less selective sample and longer exposures, multiple extreme Total SSQ scores over 100 were recorded after both training and evaluation, and the top 10% of scores were all above 60. Further, 39 out of 102 participants, or more than 38% of the sample, reported Total SSQ scores after the T2 evaluation that were greater than 20 (median SSQ after T2 = 18.7). After the T1 initial training exposure, which lasted longer than the T2 evaluation and was followed by a break before participants conducted the T2 evaluation, the percentage of scores above 20 was even higher, i.e., 43% (median T1 SSQ = 14.96).

Further, when comparing the VR and PC implementations of the external preflight inspection training software in the Phase 2 study, we observed not only statistically significantly higher SSQ scores in the VR group, the average SSQ score for the VR group was also approximately twice as high as the average in the PC group ($M_{\text{VR}} = 35.65$; $M_{\text{PC}} = 17.3$). Further, in absolute terms, the average SSQ Total scores in the VR group were, regardless of whether at T1 or T2, well above 20 ($M_{\text{VR T1(weighted)}} = 35.24$; $M_{\text{VR T2(weighted)}} = 36.06$), whereas the averages in the PC never exceeded scores of 20 ($M_{\text{PC T1(weighted)}} = 17.46$; $M_{\text{PC T2(weighted)}} = 17.14$).

Although care must be taken in applying and interpreting absolute SSQ scores (see, in particular, Bimberg, Weissker, & Kulik, 2020; Brown, Spronck, & Powell, 2022), patterns of results like the ones we saw in our studies are quite common; a recent review paper (Kim, Kim, Chung, et al., 2021) pointed to reports of 22% to 80% of participants experiencing cybersickness during or after VR exposure, and similar statistics abound both in the scientific/technical and popular literature.

Finding 33: Training organizations should be aware that cybersickness is a frequent and common occurrence with XR systems, and they should have steps in place to minimize, assess, and respond to cybersickness.

In the following, we discuss some of the common factors affecting the incidence and severity of cybersickness, and what steps organizations can take to address cybersickness. While not an exhaustive review, the focus in what follows is on identifying major contributors to cybersickness and providing practitioners with suggestions for actionable and achievable steps.

Factors affecting cybersickness. Three major groupings of influence factors, alone and in combination, affect the incidence and severity of cybersickness. Broadly, they relate to (a) the *XR technology* used to create (b) the *XR experience*, both of which interact with (c) the *individual* who experiences XR. For example, the display and computing hardware (e.g., display resolution, field of view, refresh rate, lag time) interacts with characteristics of the XR experience (such as exposure time, degree and type of self-motion in the environment, type of task, mapping of motions in the physical world to representations in the virtual world) and the individual's trait and state susceptibility to cybersickness. At low exposure times, even sub optimally designed experiences displayed on poor displays may be tolerable for most participants. In fact, reports of cybersickness are rare for exposures under 5 minutes duration; reports of substantial cybersickness typically accompany only experiences that last 15 minutes or longer. Finally, at long exposure times (over 30 minutes), with the "right" combination of technology and experiences, it is possible to create XR that will make the vast majority of users sick.

Finding 34: The incidence and severity of cybersickness is determined by the combination of XR technology with the XR experience and with the individual who uses XR. Examples of variables in each group of factors are shown in Table 10.

See Tables 10 and 11 on the subsequent pages.

Finding 35: Although there are general principles that can reduce the likelihood, incidence, and magnitude of cybersickness (see Table 11), it is the impact of their *unique combination* that is difficult to model or predict; thus, user testing is critical. This uncertainty also makes it difficult to provide minimum specifications that alone, or in combination, can ensure reduction in cybersickness.

Table 10

Selected Factors Affecting Cybersickness

1. **XR Technology**:
    1. Field of view: All else being equal, displays with larger horizontal fields of view typically increase cybersickness
    2. Display resolution, lag, flicker, and refresh rate: Low-resolution displays that have flicker and/or experience lags (in refresh rate or updates of dynamic scene changes, including motion-to-photon latency) increase the likelihood of cybersickness.
    3. Match of HMD to Inter Pupillary Distance (IPD): Mismatches increase cybersickness
    4. Match with other sensory modalities: Mismatches increase cybersickness; matching cues across modalities reduce cybersickness

2. **XR Experience**:
    1. Passive vs. active interaction: Passive XR experiences, especially with abrupt self-motion, increase cybersickness
    2. Dynamic vs. static visual content: Dynamic scene elements increase cybersickness
    3. Certain display elements: Floating objects, occlusions, and other conflicting depth cues (e.g., red lettering on a blue background), blurry display elements, display elements with high spatial frequency (e.g., striped surfaces), among others, increase cybersickness. Rest frames [e.g., nose bridge, googles], flatter textures, and fiduciary markers in the environment reduce cybersickness
    4. Ambulatory vs. standing vs. at rest/sitting: Ambulatory or standing use increases cybersickness
    5. Exposure time (acute): Longer exposures increase cybersickness
    6. Exposure time (cumulative): Unclear – most argue that cumulative exposure time reduces cybersickness through adaptation, but others have found that cybersickness monotonically increases
    7. Rest between exposures: Longer rest periods between exposures reduce cybersickness

3. **The person**:
    1. Health, motion sickness history, postural stability, mental rotation ability, and field dependence: Declines in health increase susceptibility to cybersickness
    2. XR experience and prior exposure (cumulative): Most likely, on average, more experienced participants will show lower susceptibility; but it is not clear whether through self-selection or through adaptation
    3. Age: On average, older participants are more susceptible to cybersickness, but this may be related more to health declines with age than age itself
    4. Biological sex: On average, biological females are more susceptible to cybersickness, but this may be related more to factors such as IPD mismatch, lack of experience/exposure, and other third variables, than biological sex itself

Table 11

General Principles for Reducing the Likelihood and Severity of Cybersickness

1. **XR technology**:
    1. Restrict display field of view
    2. Match hardware capabilities to graphics processing
    3. Increase display resolution
    4. Reduce display lags
    5. Reduce drift
    6. Match and synchronize outputs across sensory modalities (visual, auditory, haptic)
    7. Provide adjustable HMDs to account for Inter-Pupillary Distance (IPD)

2. **XR experience**:
    1. Avoid abrupt self-motion in the XR environment
    2. Consider sitting administration
    3. Introduce rest frames/reference markers/body markers
    4. Reduce occlusions from floating objects
    5. Reduce vection, e.g., through viewpoint snapping, when teleporting or when using external controls to change viewpoints
    6. Provide cues across multiple sensory modalities

1. **Individual**:
    1. Check for motion sickness history and current health
    2. Respect the individual
    3. Offer familiarization and adaptation options
    4. Keep exposure times relatively short
    5. Offer breaks before subsequent exposures
    6. Use a well-ventilated room
    7. Interrupt/discontinue the experience if cybersickness occurs and offer rest, ideally in fresh air or at least with an environmental view (window)
    8. Consider/provide alternative ways for training/qualification

6.0. References

Abich IV, J., Murphy, J., Eudy, M., & Killilea, J. P. (2019). Considerations for training evaluations of emerging technologies [Conference proceedings]. *MODSIM World 2019*. https://www.modsimworld.org/papers/2019/MODSIM_2019_paper_26.pdf

Abbott, K. & Donovan, C. (2020, November). Automated systems and manual flight operations: Finding a balance [Conference presentation]. *Global Aviation Training Summit—Virtual (Session 7: Pilot – Airmanship)*. https://www.gats-event.com/conference/abbott-donovan

Air Carrier Training Aviation Rulemaking Committee [ACT ARC]. (2020). Recommendation 20-11: Terms and Definitions Related to eLearning. Federal Aviation Administration Flight Standards Service.

Air Carrier Training Aviation Rulemaking Committee [ACT ARC]. (2021a). Recommendation 21-10: Flight Standardization Board Training Differences & Device Levels Requirements. Federal Aviation Administration Flight Standards Service.

Air Carrier Training Aviation Rulemaking Committee [ACT ARC]. (2021b). Recommendation 21-11: eLearning Guidance and Systematic Design Methodologies. Federal Aviation Administration Flight Standards Service.

Arce, T., & McMullen, K. (2021). The Corsi block-tapping test: Evaluating methodological practices with an eye towards modern digital frameworks. *Computers in Human Behavior Reports*, 4, 100099. https://doi.org/10.1016/j.chbr.2021.100099

Azevedo, Roger & Gasevic, Dragan. (2019). Analyzing Multimodal Multichannel Data about Self-Regulated Learning with Advanced Learning Technologies: Issues and Challenges. *Computers in Human Behavior*. 96. 207-210. 10.1016/j.chb.2019.03.025.

Bedwell, W. L., & Salas, E. (2010). Computer-based training: capitalizing on lessons learned. *International Journal of Training and Development*, *14(3)*, 239-249. https://doi.org/10.1111/j.1468-2419.2010.00355.x

Bell, B. S., & Moore, O. A. (2018). Learning, training and development in organizations: Emerging trends, recent advances and future directions. In D. S. Ones, N. Anderson, C. Viswesvaran, & H. K. Sinangil (Eds.), *The Handbook of Industrial, Work & Organizational Psychology* (2nd ed., pp. 215-233). Thousand Oaks, CA: Sage. https://core.ac.uk/download/pdf/219376768.pdf

Bimberg, Pauline & Weissker, Tim & Kulik, Alexander. (2020). On the Usage of the Simulator Sickness Questionnaire for Virtual Reality Research. 464-467. 10.1109/VRW50115.2020.00098.

Brown, P., Spronck P., & Powell W. (2022). The simulator sickness questionnaire, and the erroneous zero baseline assumption. *Frontiers in Virtual Reality*, *3*, DOI:10.3389/frvir.2022.945800.

Brusso, Robbie & Cigularov, Konstantin & Callan, Rachel. (2014). Longitudinal Research in Occupational Stress. 10.1002/9781118539415.wbwell034.

Casner, S. M., Geven, R. W., Recker, M. P., & Schooler, J. W. (2014). The retention of manual flying skills in the automated cockpit. *Human Factors*, *56(8)*, 1506–1516. https://doi.org/10.1177/0018720814535628

Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development*, *42(2)*, 21-29. https://doi.org/10.1007/BF02299088

Collins Aerospace. (2023). *Coalescence. Collins Aerospace.* *https://www.collinsaerospace.com/what-we-do/industries/military-and-defense/simulation-and-training/training-systems/coalescence*

Corsi, P. (1972) Memory and the Medial Temporal Region of the Brain. Doctoral Thesis in Philosophy, McGill University, Montreal.

Cruit, J. K., Bockelman, P., Hancock, P. (2020, March 4). Redefining resiliency using principles of social cognitive theories to address the gap in training and assessing resiliency for air carrier operations [Conference presentation]. *National Training Aircraft Symposium: Concurrent Session* 9B: The Changing Landscape of Flight Training, 62. Daytona Beach, FL. https://commons.erau.edu/ntas/2020/presentations/62

Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology*, *19(2)*, 272-309.

Dalgarno, B., & Lee, M. J. (2010). What are the learning affordances of 3-D virtual environments? *British journal of educational technology*, *41(1)*, 10-32.

Degani, A., & Wiener, E. L. (1997). Procedures in complex systems: The airline cockpit. IEEE Transactions on Systems, Man, and Cybernetics-Part A: *Systems and Humans*, *27(3)*, 302-312. https://doi.org/10.1109/3468.568739

Dinçer, N. (2023). Elevating Aviation Education: A Comprehensive Examination of Technology's Role in Modern Flight Training. *Journal of Aviation*,7 (2), 317- 323. DOI:10.30518/jav.1279718

Dunne, R., Schatz, S., Fiore, S. M., Marin, G., & Nicholson, D. (2010, September). Scenario-based training: scenario complexity. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *54(27)*, 2238-2242. Sage CA: Los Angeles, CA: SAGE Publications. https://doi.org/10.1177%2F154193121005402704

Effectiveness of Knowledge Training Workgroup. (2020, January). Terms of Reference. Federal Aviation Administration, Flight Standards Service, Air Carrier Training Aviation Rulemaking Committee. https://www.faa.gov/sites/faa.gov/files/about/office_org/headquarters_offices/avs/ACT_ARC_Rec_21-11.pdf

Elias, B. (2019, October 3). Cockpit automation, flight systems complexity, and aircraft certification: Background and issues for Congress (CRS Report No. R45939). U.S. Congressional Research Service. https://crsreports.congress.gov/product/pdf/R/R45939/2

Etherington, T., Kramer, L. J., Young, S. D., & Daniels, T. S. (2020, January). Information management to mitigate loss of control inflight airline accidents. *American Institute of Aeronautics and Astronautics: Scitech 2020 Forum*, 0054. https://doi.org/10.2514/6.2020-0054

Federal Aviation Administration [FAA]. (2017b). *Advanced Qualification Program* (FAA-AC-120-54A). Washington, D.C.: U.S. Department of Transportation. https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_120-54A_CHG_1.pdf

Faulconer, E. K., & Gruss, A. B. (2018). A review to weigh the pros and cons of online, remote, and distance science laboratory experiences. *International Review of Research in Open and Distributed Learning*, 19(2). https://doi.org/10.19173/irrodl.v19i2.3386

Filburn, T. (2020) Flight system sensor failure. In T. Filburn (Ed.), Commercial Aviation in the Jet Era and the Systems that Make it Possible (pp. 169-179). Springer, Cham. https://doi.org/10.1007/978-3-030-20111-1_14

Fortin, A., Viger, C., Deslandes, M., Callimaci, A., & Desforges, P. (2019). Accounting students' choice of blended learning format and its impact on performance and satisfaction. *Accounting Education*, *28(4)*, 353-383. https://doi.org/10.1080/09639284.2019.1586553

Graesser, A. C. (2013). Evolution of advanced learning technologies in the 21st century. *Theory Into Practice*, *52(1)*, 93-101. https://doi.org/10.1080/00405841.2013.795446

Hoffman, R., Feltovich, P., Fiore, S., Klein, G., Missildine, W., & DiBello, L. (2010). Accelerated proficiency and facilitated retention: Recommendations based on an integration of research and findings from a working meeting. *Florida Institute for Human and Machine Cognition, Inc.* Pensacola, FL.

Human Factors & Aviation Safety. Hearing on The Boeing 737 MAX: Examining the Federal Aviation Administration's Oversight of the Aircraft's Certification before The House Committee on Transportation & Infrastructure, House of Representatives, 116th Cong. 1-17 (2019, December 11) (testimony of Mica Endsley). https://transportation.house.gov/imo/media/doc/Endsley%20Testimony.pdf

Iván Aguilar Reyes, C., Wozniak, D., Ham, A., & Zahabi, M. (2022). An adaptive virtual reality- based training system for pilots. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 66, No. 1, pp. 1962-1966). Sage CA: Los Angeles, CA: SAGE Publications.

Jentsch, F., & Curtis, M. (2017). Simulation in aviation training. Routledge

Johnson-Glenberg, M. C. (2018). Immersive VR and education: Embodied design principles that include gesture and hand controls. *Frontiers in Robotics and AI*, *5*. https://doi.org/10.3389/frobt.2018.00081

Kaplan, A., Cruit, J., Endsley, M., Beers, S., Sawyer, B., & Hancock, P. (2020). The Effects of Virtual Reality, Augmented Reality, and Mixed Reality as Training Enhancement Methods: A Meta-Analysis. Human Factors. DOI: 10.1177/0018720820904229.

Kauchak, M. (2022). AR/MR/VR/XR as training disruptors. *Civil Aviation Training Magazine*, *33(1)*, 22-30. https://www.halldale.com/cat1_2022_digital

Kim, H., Kim, D.J., Chung, W.H. *et al.* (2021). Clinical predictors of cybersickness in virtual reality (VR) among highly stressed people. *Sci Rep* 11, 12139 (2021). https://doi.org/10.1038/s41598-021-91573-w

Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum & Michael G. Lilienthal (1993) Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness, The International Journal of Aviation Psychology, 3:3, 203-220, DOI: 10.1207/s15327108ijap0303_3

Kessels RP, van den Berg E, Ruis C, Brands AM. The backward span of the Corsi Block-Tapping Task and its association with the WAIS-III Digit Span. Assessment. (2008 Dec).15(4):426-34. DOI: 10.1177/1073191108315611. Epub 2008 May 15. PMID: 18483192

Kim, J. W. (2008). Procedural skills: From learning to forgetting (Order No. 3336051) [Dissertation manuscript, Pennsylvania State University]. ProQuest Dissertations & Theses A&I; ProQuest Dissertations & Theses Global. (304496168). https://www.proquest.com/dissertations-theses/procedural-skills-learning-forgetting/docview/304496168/se-2

Kim, C., Sung, J., Han, J. Y., Jee, S., Lee, J. W., Lee, J. H., & Kwon, O. P. (2021). Current status of cardiac rehabilitation in the regional cardiocerebrovascular centers in Korea. *Journal of Clinical Medicine*, *10(21*), 5079.

Koedinger, Kenneth & Corbett, Albert & Perfetti, Charles. (2012). The Knowledge-Learning-Instruction (KLI) Framework: Toward Bridging the Science-Practice Chasm to Enhance Robust Student Learning. *Cognitive science. 36*. 757-98. 10.1111/j.1551-6709.2012.01245.x.

Lapsley, R., Kulik, B., Moody, R., & Arbaugh, J. B. (2008). Is identical really identical? An investigation of equivalency theory and online learning. *Journal of Educators Online, 5(1)*, n1. http://evidence.thinkportal.org/handle/123456789/17187 http://dx.doi.org/10.9743/JEO.2008.1.3

Lawrynczyk, A. (2018). Exploring virtual reality flight training as a viable alternative to traditional simulator flight training. *Carleton University Research Virtual Environment.* https://curve.carleton.ca/8befeba3-7102-4851-a690-7a513e0f8343

Lewis, J., & Livingston, J. (2018). Pilot training next: Breaking institutional paradigms using student-centered multimodal learning [Paper presentation]. *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2018,* Orlando, FL, United States.

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication, 3(2).* DOI: 10.1111/j.1083-6101.1997.tb00072.x

Long, Yanjin & Aleven, Vincent. (2013). Supporting Students' Self-Regulated Learning with an Open Learner Model in a Linear Equation Tutor. 7926. 10.1007/978-3-642-39112-5_23.

Maher, M. L., Zhang, D., & Ennis, G. (2000). Virtual reality in landscape design. Landscape and Urban Planning, *52(1)*, 167-178.

Makransky, G., & Petersen, G. B. (2021). The cognitive-affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review, 1-22.*

Makransky, G., Andreasen, N. K., Baceviciute, S., & Mayer, R. E. (2020a). Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. *Journal of Educational Psychology.* https://doi.org/10.1037/edu0000473 .

Matton, N., Vrignaud, C., Rouillard, Y., Lemarié, J. (2018). Learning flight procedures by enacting and receiving feedback. *Applied Ergonomics, 70,* 253-259. https://doi.org/10.1016/j.apergo.2018.03.009

McGowin, Greg & Fiore, Stephen & Oden, Kevin. (2021). Learning Affordances: Theoretical Considerations for Design of Immersive Virtual Reality in Training and Education. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 65.* 883-887. https://doi.org/10.1177/1071181321651293

McGowin, G., Fiore, S.M., Oden, K. (2022). Me, Myself, and the (Virtual) World: A Review of Learning Research in 4E Cognition and Immersive Virtual Reality. In: Chen, J.Y.C., Fragomeni, G. (eds) *Virtual, Augmented and Mixed Reality: Applications in Education, Aviation and Industry. HCII 2022*. *Lecture Notes in Computer Science*, vol. 13318. Springer. https://doi.org/10.1007/978-3-031-06015-1_5

Miller, H. L., & Bugnariu, N. L. (2016). Level of immersion in virtual environments impacts the ability to assess and teach social skills in autism spectrum disorder. C*yberpsychology, Behavior, and Social Networking*, *19(4)*, 246-256. https://doi.org/10.1089/cyber.2014.0682

Muñoz-Marrón, D., Gil, F., & Lanero, A. (2018). Are crews empowered with all the resources needed to successfully address an inflight emergency? Checklists, a necessary but inefficient tool. *Aviation*, *22(3)*, 93-101. https://doi.org/10.3846/aviation.2018.6254

National Transportation Safety Board. (2019a, September 19). Safety recommendation report: Assumptions used in the safety assessment process and the effects of multiple alters and indications on pilot performance. ASR1901. Assumptions Used in the Safety Assessment Process and the Effects of Multiple Alerts and Indications on Pilot Performance (ntsb.gov)

National Transportation Safety Board. (2019b, September 26). NTSB issues 7 safety recommendations to FAA related to ongoing Lion Air, Ethiopian Airlines crash investigations. https://www.ntsb.gov/news/press-releases/Pages/NR20190926.aspx

Nguyen, B., Sonnenfeld, N. A., Finkelstein, L., Alonso, A., Duruaku, F., Gomez, C., & Jentsch, F. (in press). Using AI tools to develop training materials for aviation: Ethical, technical, and practical concerns. *Proceedings of the Human Factors & Ergonomics Society Annual Meeting.*

Oberhauser, M., Dreyer, D., Braunstingl, R., & Koglbauer, I. (2018). What's real about virtual reality flight simulation? *Aviation Psychology and Applied Human Factors*.

Orlansky, J., Dahlman, C. J., Hammon, C. P., Metzko, J., & Taylor, H. L. (1994). The Value of Simulation for Training (No. IDA-P-2982). *Institute for Defense Analysis*. Alexanderia, VA. https://apps.dtic.mil/sti/pdfs/ADA289174.pdf

Petersen, G. B., Petkakis, G., & Makransky, G. (2022). A study of how immersion and interactivity drive VR learning. *Computers & Education*, *179*, 104429.

Peterson B, Hawke F, Spink M, Sadler S, Hawes M, Callister R, Chuter V. Biomechanical and Musculoskeletal Measurements as Risk Factors for Running-Related Injury in Non-elite Runners: A Systematic Review and Meta-analysis of Prospective Studies. *Sports Med Open.* 2022 Mar 7;8(1):38. DOI: 10.1186/s40798-022-00416-z. PMID: 35254562; PMCID: PMC8901814.

Povenmire, H. K., & Roscoe, S. N. (1973). Incremental Transfer Effectiveness of a Ground-Based General Aviation Trainer. *Human Factors, 15(6)*, 534–542. https://doi.org/10.1177/001872087301500605

Rauschnabel, Philipp & Babin, Barry & Tom Dieck, M. Claudia & Krey, Nina & Jung, Timothy. (2022). What is Augmented Reality Marketing? Its definition, complexity, and future. *Journal of Business Research. 142.* 1140-1150. 10.1016/j.jbusres.2021.12.084.

Risukhin, V. N. (2016). Integration of affordable information technology products into general aviation training and research. In 16th *AIAA Aviation Technology, Integration, and Operations Conference* (p. 3917).

Roscoe, S. N., & Bergman, C. A. (1980). Flight performance control. In S.N. Roscoe (Ed.) *Aviation Psychology* (Chapter 4, pp. 39-47). Ames: Iowa State University Press.

Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience, 6(4)*, 332-339. DOI: 1038/nrn1651

Saraçyakupoğlu, T. (2020). The Novel Manufacturing Techniques of Airborne Parts. In *Theory and Research in Engineering* (Vol. II, pp. 179–194). essay, Gece Publishing.

Sawyer, M. W., Berry, K. A., Kring, J. P., Austrian, E. M. (2018, August 14). A human-systems approach to proactively managing risk through training in an evolving aviation industry. *National Training Aircraft Symposium: Session 7 - Safety Management Systems*. Daytona Beach, FL. https://commons.erau.edu/ntas/2018/presentations/31

Schaffernak, H., Mösl, B., Vorraber, W., & Koglbauer, I. (2020). Potential augmented reality application areas for pilot education: An exploratory study. *Education Sciences*, *10(4),* 86. https://doi.org/10.3390/educsci10040086

Schaffernak, H., Moesl, B., Vorraber, W., Holy, M., Herzog, E.-M., Novak, R., & Koglbauer, I.V. (2022). Novel mixed reality use cases for pilot training. *Education Sciences*, *12,* 345. https://doi.org/10.3390/educsci12050345

Seymour, N. E., Gallagher, A. G., Roman, S. A., O'Brien, M. K., Bansal, V. K., Andersen, D. K., & Satava, R. M. (2002). Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Annals of Surgery*, *236(4),* 458-463.

Sherman, W. R., & Craig, A. B. (2003). Understanding virtual reality: Interface, application, and design. Elsevier.

Simonson, M. (1999). Equivalency theory and distance education. In K. E. Sparks and M. Simonson (Eds.), *Proceedings of Selected Research and Development Papers Presented at the National Convention of the Association for Educational Communications and Technology* (pp. 207-210). Houston, TX. https://files.eric.ed.gov/fulltext/ED436128.pdf

Slater, M. (2003). A note on presence terminology. *Presence Connect, 3(3),* 1-5.

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, *6(6),* 603-616. https://doi.org/10.1162/pres.1997.6.6.603

Sonnenfeld, N. A., Nguyen, B., Boesser, C. T., & Jentsch, F. (2021a, May). Modern training practices for flightcrew training of procedural knowledge. *International Symposium on Aviation Psychology, 21,* 305-309. https://doi.org/10.5399/osu/1148

Sonnenfeld, N. A., Nguyen, B., Boesser, C. T., Duruaku, F., & Jentsch, F. (2021b, October 21). *Considerations for Electronic & Distance Learning of Procedures in Flight Crew Training* [Unpublished technical report]. Orlando: University of Central Florida.

Sonnenfeld, N., Nguyen, B., Duruaku, F., Alonso, A., Carney, M., Jentsch, F., & Smith, E. (2023). Flightcrew Procedures Experimental Training: Simulation Testbed Design, Development, & Interactive Demonstration. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.*

Walters, W. T., & Walton, J. (2021). Efficacy of virtual reality training for pilots: a review of links between user presence, search task performance, and collaboration within virtual reality. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 65, No. 1, pp. 919-922). Sage CA: Los Angeles, CA: SAGE Publications.

Wilkinson, M., Brantley, S., & Feng, J. (2021). A Mini Review of Presence and Immersion in Virtual Reality. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *65(1),* 1099–1103. https://doi.org/10.1177/1071181321651148

Van Weelden, E., Alimardani, M., Wiltshire, T. J., & Louwerse, M. M. (2021). Advancing the adoption of virtual reality and neurotechnology to improve flight training. In *2021 IEEE 2nd International Conference on Human-Machine Systems (ICHMS)* (pp. 1-4). IEEE.

Vandierendonck, Dirk & Haynes, Clare & Borrill, Carol & Stride, Chris. (2004). Leadership Behavior and Subordinate Well-Being. *Journal of occupational health psychology. 9.* 165-75. 10.1037/1076-8998.9.2.165.

Appendix A: Supplemental Materials Furnished to the FAA

The following supplemental materials have been furnished to the FAA under the grant. The asterisk symbol ("*") indicates that it has been previously submitted to the FAA

Phase 1 Supplements

- Supplement A: Phase 1 Research Report*
  - Supplement A1: Phase 1 Research Report Final Draft*
- Supplement B: Phase 1 Contributions to ACT ARC EKT WG
  - Supplement B1: E&DL Glossary*
  - Supplement B2: Foundations of Systematic Models/Frameworks for TEE*
- Supplement C: Materials for Phase 1 Conference Proceedings
  - Supplement C1: ISAP 2021 Extended Abstract*
  - Supplement C2: ISAP 2021 Conference Proceedings Paper*
  - Supplement C3: ISAP 2021 Conference Proceedings Presentation*

Phase 2 & Phase 3 Supplements

- Supplement D: Materials for Phase 2 Simulation Development
  - Supplement D1: FlightPET User Manual v0.1.38
- Supplement E: Materials for Phase 2 Presentations and Conference Proceedings
  - Supplement E1: HF Roundtable Presentation*
  - Supplement E2: HFES 2023 Proceedings—FlightPET Testbed*
  - Supplement E3: HFES 2023 Proceedings—AI Tools*
  - Supplement E4: HFES 2023 Proceedings—TEE Frameworks*
  - Supplement E5: HFES 2023 Proceedings—Suitability of VR*
  - Supplement E6: I/ITSEC Proceedings Paper—Episodic Memory*
  - Supplement E7: NEAT Presentation*