DOT/FAA/TC-21/1

Federal Aviation Administration
William J. Hughes Technical Center
Aviation Research Division
Atlantic City International Airport
New Jersey 08405

Optimization of Air Traffic Control Information Presentation (OAIP) in the En Route Environment: Baseline Simulation

January 2021

Final report

U.S. Department of Transportation
Federal Aviation Administration

NOTICE

This report is available at the Federal Aviation Administration William J. Hughes Technical Center's Full-Text Technical Reports page: www.faa.gov/about/office_org/headquarters_offices/ang/library/ in Adobe Acrobat portable document format (PDF).

Technical Report Documentation Page

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/TC-21/1		

4. Title and Subtitle	5. Report Date
Optimization of Air Traffic Control Information Presentation (OAIP) in the En Route Environment: Baseline Simulation	January 2021
	6. Performing Organization Code ANG-E25
7. Author(s) Carolina M. Zingale, Ph.D., FAA Human Factors Branch; Brion Woroch, Ph.D., SAIC; Matthew Dworsky, SAIC; Ben Willems, FAA Program Management Office	8. Performing Organization Report No. DOT/FAA/TC-21/1
9. Performing Organization Name and Address Federal Aviation Administration Human Factors Branch William J. Hughes Technical Center Atlantic City International Airport, NJ 08405	10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address Federal Aviation Administration Human Factors Research and Engineering Group 800 Independence Avenue, S.W. Washington, DC, 20591	13. Type of Report and Period Covered 14. Sponsoring Agency Code ANG-C1

15. Supplementary Notes

16. Abstract

The Optimization of Air Traffic Control Information Presentation (OAIP) project investigates the effects of co-locating new tools and capabilities onto the existing air traffic control systems. Software developers at the FAA Research Development and Human Factors Laboratory at the William J. Hughes Technical Center developed a platform to simulate the existing en route air traffic system and several planned new tools and capabilities that included Conflict Probe on the Radar-side (R-side) display, Airborne Reroute (ABRR), Time-based Flow Management (TBFM), electronic Controller-to Controller coordination (CTCC), Data Communications (Data Comm), and a 43-inch display at the R-side position. Sixteen current en route air traffic controllers participated in a simulation to utilize the tools in a realistic air traffic environment. Half of the participants had 15 or more years of en route air traffic control experience (High Experience) and half had 5 or fewer years of experience (Low Experience). The participants worked moderate and busy traffic level scenarios either as R-side controllers alone or with a Radar Associate (RA). We recorded all system activity, including participant entries and communications with simulation pilots, and gathered subjective workload ratings and comments. We monitored R-side participant eye movements and collected encephalogram (EEG) recordings to measure brain region activity. Air traffic metrics (e.g., handoffs, airspace deviations) indicated that R-side participants had more difficulty managing traffic in the higher-traffic-level scenarios or when working alone. We also found larger pupil diameters—an indicator of workload—in high-traffic-level scenarios. EEG analyses revealed brain activity in areas associated with attention, decision-making, change detection, semantic meaning extraction, object processing, and response planning. The participants made some use of the new tools, but also provided feedback about difficulties using them. These difficulties included confusing symbology, display clutter, confusing use of color, and cumbersome or confusing command entries. The participants in the High Experience group reported more difficulties than those in the Low Experience group. The participants reported that working with an RA was important, especially in busy scenarios. Having an RA increased overall tool use. However, adding tools will require training and a clear designation of roles and responsibilities to establish effective teamwork between R-side and RA.

17. Key Words	18. Distribution Statement
Air Traffic Control En Route Data Communications Conflict Probe Airborne Reroute Time-based Flow Management	This document is available to the U.S. public through the National Technical Information Service (NTIS), Springfield, Virginia 22161. This document is also available from the Federal Aviation Administration William J. Hughes Technical Center at www.faa.gov/about/office_org/headquarters_offices/ang/library

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	127	

Acknowledgements

We thank the ANG-E142 and SAIC laboratory team who developed and implemented the en route system simulator and the new tools and capabilities needed for this study and who ensured that all systems and data recording capabilities were operating as needed.

We thank the Target Generation Facility staff from ANG-161 and JVN Communications who ensured that aircraft performance and simulation pilot workstations functioned as needed, and we thank the simulation pilots from ANG-E161 and EIT, Inc. who managed aircraft at the workstations and handled communications with controllers during simulation shakedown and data collection.

We thank Adam Rhodes, NATCA National NextGen Representative, for working diligently to coordinate air traffic controller participation for this simulation. We thank the 16 en route air traffic controllers from facilities throughout the US who participated in the study.

We thank Nelson Brown, ANG-E25, and Laura Hamann, ANG-142, for extracting and consolidating the aircraft and system data from the simulation for analysis. We thank Albert Macias, ANG-E141, for extracting and consolidating the eye movement data. We thank Olivia McGough, student intern from Rowan University Psychology Department, for working with us to organize and analyze data.

Contents

Figures

Tables

Acronyms

Acronym	Definition
ABRR	Airborne Reroute
ACL	Aircraft List
ADS-B	Automatic Dependent Surveillance - Broadcast
AMS	Acquisition Management System
ANODE	Analysis of Deviance
ANOVA	Analysis of Variance
ARTCC	Air Route Traffic Control Center
ATC	Air Traffic Control
CHI	Computer-Human Interface
CID	Computer Identification Number
CONUS	Continental United States
CP	Conflict Probe
CPC	Certified Professional Controller
CTCC	Controller-to-Controller Coordination
Data Comm	Data Communications
DC-FS	Data Communication – Full Services version
DCT	Display Countdown Timer
DESIREE	Distributed Environment for Simulation Rapid Engineering and Experimentation
DOT	Department of Transportation
EDDS	En Route Data Distribution System
EDSM	En Route Display System Management
EEG	Electroencephalogram
ER2	Experiment Room #2
ER3	Experiment Room #3
ERAM	En Route Automation Modernization
FAA	Federal Aviation Administration
FDB	Full Data Block
fERP	Fixation Event Related Potential
FL	Flight Level
GIM-S	Ground Interval Management- Speed

GLM	General Linear Model
HF	Human Factors
HITL	Human-in-the-Loop
IC	Independent Component
ICA	Independent Component Analysis
JEDI	Joint En Route Decision Support System Infrastructure
KIAD	International designation for Dulles Airport
KPHL	International designation for Philadelphia Airport
LDB	Limited Data Block
LOA	Letter of Agreement
LOS	Loss of Separation
M	Mean
MAP	Monitor Alert Parameter
MRP	Meter Reference Point
NATCA	National Air Traffic Control Association.
OAIP	Optimization of Air Traffic Control Information Presentations
PEQ	Post Experiment Questionnaire
PMO	Project Management Office
POG	Point of Gaze
PSQ	Post Scenario Questionnaire
PTT	Push-to-Talk
R-side	Radar side
RA	Radar Associate
RDHFL	Research Development Human Factors Laboratory
RVSM	Reduced Vertical Separation Minimum
SD	Standard Deviation
SME	Subject Matter Expert
SOP	Standard Operating Procedure
STA	Schedule Time of Arrival
STARS	Standardized Terminal Automation Replacement System
SAA	Special Activity Airspace
TAMR	Terminal Automation Modernization Replacement
TBFM	Time-Based Flow Management
TFM	Traffic Flow Management
TGF	Target Generation Facility

TOD	Top of Descent
URET	User Request Evaluation Tool
VCI	Voice Communication Indicator
WAK	Workload Assessment Keyboard
WJHTC	William J. Hughes Technical Center
ZNY	Identification for New York Center ARTCC

Executive Summary

The Optimization of Air Traffic Control (ATC) Information Presentation (OAIP) project is a multi-year, multi-phase effort in which researchers will simulate the effects of co-locating new tools and capabilities onto the en route air traffic control system. The goal of the current project is to conduct a human-in-the-loop simulation with current air traffic controllers to evaluate the effectiveness and safety of the computer-human interface (CHI) design decisions made for these tools and to identify any human factors issues associated with them.

The project uses the resources and capabilities of the William J. Hughes Technical Center (WJHTC) Research Development Human Factors Laboratory (RDHFL) to present proposed ATC system tools and capabilities on a single air traffic simulation platform. This report summarizes the results of a baseline simulation that investigated the effect of Conflict Probe on the Radar (R-side) display, Airborne Reroute (ABRR), Time-Based Flow Management (TBFM), electronic Controller-to-Controller Communication (CTCC), Data Communications (Data Comm), and a 43-inch monitor at the R-side position.

Sixteen current certified professional controllers (CPC) participated in the 8-day simulation. We designed the experiment to include eight participants with 15 or more years of experience in the en route environment and eight participants with five or fewer years of experience. This enabled us to evaluate whether ATC experience affected the participants' performance and their subjective impressions of the tools. We recorded all system activity, including participant interactions with the simulator and communications with simulation pilots. We monitored and recorded participant eye movements and brain activity. We gathered subjective workload ratings throughout the simulation and on questionnaires. The participants provided additional feedback and reactions to the tools on the questionnaires and during a final debrief.

We included two traffic scenarios, one with a moderate level of traffic and the other with a higher level of traffic. Each participant completed a total of 16 test scenarios. They worked half of the scenarios as R-side controllers alone and the other half in R-side/Radar Associate (RA) teams. We analyzed the data to determine whether experience level, scenario type, and team configuration affected their workload, the way they interacted with the system, their use of the new tools and capabilities, and their ability to control traffic safely and efficiently.

The participants differed in their reactions to the new tools and capabilities based on their experience level. The Low Experience participants generally reported that the information provided by the tools had a more positive effect on their performance, control of traffic, situation monitoring, management of sector resources, and rerouting and evaluating flight plans than did

the High experience participants. However, the Low Experience participants also reported greater effort and frustration and greater physical and temporal demand than did the High Experience participants. The High Experience participants also rated some of the symbology to be more confusing than the Low Experience participants and reported that some functions required more reliance on working memory. These differences may influence the extent to which individuals make use of the tools and could negatively affect R-side/RA team coordination.

We found, as expected, that the participants managed more aircraft and made more voice transmissions in the busier traffic scenarios than the moderate traffic level scenarios. The participants also typically reported higher workload when they worked the higher-traffic-level scenarios and when they worked alone at the R-side. Eye movement and EEG data provided two objective measures of workload and cognitive demand. We expected that higher demand may result when the participants processed additional data block information provided by the new tools. We expected to find a relationship between larger pupil diameter—a measure of workload—and data block complexity, but we did not find a significant correlation. However, we did find larger pupil diameters in the busy traffic level scenarios overall.

We analyzed the EEG data to evaluate participant brain activity. Our EEG analyses identified activity in brain regions involved in a variety of cognitive processes, such as: attention, decision-making, change detection, semantic meaning extraction, object processing, and response planning—areas that would be expected given the complex nature of air traffic control. We created Fixation Event Related Potentials (fERPs) that were time-locked to participant fixations on data blocks to examine whether fERPs changed with data block complexity, but we did not find significant results. We did find some effects of our experimental conditions on the fERPs that can provide a baseline for subsequent studies. We believe that our approach to measuring and analyzing EEG data will be useful in future air traffic control simulations.

We evaluated the participants' use of the new tools and capabilities by examining the number of participant interactions associated with them—use of the trackball to select a display element or use of the keyboard for data entry. The R-side participants interacted more with the new tools when they worked alone than when they worked with an RA, though the R-side/RA teams used the new tools more overall.

Individual tool use varied, as follows:

- <u>Conflict Probe</u>: The participants did not make much use of Conflict Probe on the R-side. Half of the participants did not use the tool at all and many reported that displaying the conflicting routes caused display clutter.

- TBFM: The participants reported that although the TBFM concept was useful, the tool was not well suited to the small sector used in this study. The majority of interactions (70%) were rejections of the TBFM system recommendations. The participants also indicated some confusion about the symbol ("C") used to designate a speed advisory.

- Data Comm: The participants found Data Comm very useful, especially for uplinking route clearances to pilots. They interacted more with the tool when they worked the R-side position alone and when they worked moderate traffic level scenarios. The participants reported that they resorted to "talking and turning" when they were busy. They also found that the Data Comm menu structures could be confusing.

- ABRR: The participants found ABRR most helpful when used in conjunction with a Data Comm uplink. But, some participants indicated that they were confused by the color-coding used.

- CTCC: The participants had mixed reactions to CTCC. Some participants reported that the symbology—green triangle—was not salient. Several participants reported that a coordination call would be faster because of the time required to navigate CTCC menus and clear automatically-filled fields. We also observed that the participants had difficulty selecting the correct trackball button to initiate the CTCC function.

In summary, the OAIP project developed a laboratory simulation platform for investigating the effect of co-locating new tools and capabilities on the existing air traffic system. The project also integrated novel EEG measurement techniques to evaluate controller brain region activity. These techniques will be useful in future studies. The simulation identified several important human factors issues associated with new workstation tools and capabilities, including: display clutter, confusing symbology and use of color, and confusing command entries. The simulation also highlighted the need to clearly designate the roles and responsibilities of the R-side and RA controllers as new tools are added to the system to ensure that controllers can work effectively as a team. The participants differed in their reactions to the new tools and capabilities based on their experience levels, with low experience controllers more receptive to the use and benefits of the new tools and high experience controllers reporting more difficulty interpreting the display symbology introduced by the tools. FAA system acquisition teams should address these human factors issues to improve the usefulness of the new tools and to streamline their integration into the operational environment.

1 Introduction

The Optimization of Air Traffic Control (ATC) Information Presentation (OAIP) project investigates the impact of adding multiple new tools and capabilities to the existing air traffic control systems. Due to program funding and FAA's program management approach, which must comply with the Acquisition Management System (AMS), new controller workstation functions are designed and developed in piecemeal fashion, independent of one another. Each new function implements design and interaction strategies and uses symbols and colors that differ from the others and from those used in the existing controller computer-human interface (CHI) in the en route air traffic system. Such design inconsistencies can cause confusion and delay the controller's ability to respond quickly, decisively, and accurately. Adding new tools to the existing system may also cause display clutter. Unless they are resolved, these human factors (HF) issues will affect National Airspace System (NAS) safety and efficiency, as air traffic controllers encounter difficulties using the new tools.

The OAIP project is a multi-year effort that seeks to identify and propose mitigations to identified HF issues. The first segment of the project focuses on the en route domain. The new capabilities for the en route environment include the addition of a 43" display at the radar position (R-side), the implementation of Conflict Probe at the R-side position, Data Communications (Data Comm), Time-Based Flow Management (TBFM), Airborne Reroute (ABRR), and electronic controller-to-controller communication (CTCC). Phase 1 of the en route OAIP project involved a desktop review of the capabilities by researchers to identify potential HF issues that may arise from adding the new tools and capabilities to the current system (Willems & Dworsky, 2018). The reviewers followed a human-centered design methodology to capture and categorize issues and assign severity scores to the items. The HF reviewers found several issues that could negatively affect system use, including design inconsistencies, convention violations, a need for users to rely heavily on working memory, interruptions to automatic processes, annotation complexities, and display clutter.

This document describes Phase 2 of the en route OAIP project that involved a human-in-the-loop (HITL) simulation. We co-located the new tools and capabilities onto a simulated en route air traffic system to evaluate the effect on air traffic controller performance and workload and to validate the issues identified in the HF review. The researchers designed the simulation to provide the tools and capabilities as they were planned for future use in the en route ATC environment and to provide a baseline to which a modified system configuration could later be compared in a subsequent project phase. This document summarizes the design, conduct, and findings of this baseline simulation.

2 Purpose

The purpose of the simulation was to identify HF issues that may arise from co-locating multiple new tools and capabilities in the en route ATC environment. We recruited current en route air traffic controllers to participate in the simulation. The research team trained the participants on the use of the new tools and capabilities and then the participants completed test scenarios using them. We recorded system data during the simulation to evaluate the participants' interaction with the system and their ability to manage traffic safely and efficiently. We also obtained subjective reports from the participants about their workload and gathered their feedback regarding the new tools and capabilities. We monitored and recorded participant eye movements and electroencephalographic (EEG) data during the test scenarios to obtain additional measures of cognitive and visual workload. We used the results of the simulation to identify and prioritize HF issues.

3 Methodology

Researchers at the Federal Aviation Administration (FAA) William J. Hughes Technical Center (WJHTC) Research Development and Human Factors Laboratory (RDHFL) developed and conducted the simulation. Software developers at the RDHFL updated the existing simulated en route air traffic control system to enable the new capabilities and connected the simulated system to other required tools as described below.

In coordination with the Program Management Office (PMO) and National Air Traffic Controllers Association (NATCA), we recruited sixteen en route certified professional controllers (CPCs) from Air Route Traffic Control Centers (ARTCCs) throughout the continental United States (CONUS). The participants completed the simulation in groups of four at a time over a two-week period. The participants traveled in on Monday of the first week and traveled out on Friday of the second week. We provided training and familiarization during the first week and conducted the test scenarios during the second week. We conducted data collection between October 2019 and February of 2020.

3.1 Participants

Sixteen CPCs from nine ARTCCs throughout the CONUS participated in the simulation, half of whom had 15 or more years of CPC experience (High Experience) and half of whom had five or fewer years of CPC experience (Low Experience). We wanted to determine whether controller experience affected the way in which the participants interacted with the system and the new tools and capabilities and whether their subjective impressions of them differed. Due to recruitment difficulties, one of the participants included in the High Experience group had only 8.5 years of CPC experience. However, because he had over 20 years of overall air traffic control experience, we included him in this group to ensure we had the total number of participants required to complete the schedule.

The participants completed the simulation in groups of four at a time, 2 from the Low Experience group (L) and 2 from the High Experience group (H). They worked as either an R-side controller alone or as part of an R-side/RA team (with RA). We paired the participants by experience level (L-L; H-H) so that the individuals in each team would be as similar to one another as possible. Due to time constraints, we were unable to test all possible team configurations that would have included pairs of participants with different levels of experience. As time allowed, however, we included a “mixed” team configuration scenario on testing days, but did not include these scenarios in the primary data analyses.

After listening to the initial briefing describing the simulation, the participants read and signed the Informed Consent Statements (Appendix A) that provided a summary of the simulation purpose and the participants’ rights and responsibilities. Next, we obtained demographic data from the participants via a Background Questionnaire (Appendix B). All of the participants in the simulation were male with median age of 33. The median age of the participants in the Low Experience group was 30, and the median age of the participants in the High Experience group was 43. Including military experience, the participants in the Low Experience group had a median of 6 years’ experience as air traffic controllers, and the participants in the High Experience group had a median of 24 years’ experience.

All of the participants were current in the en route environment. Five of the participants also had experience in the terminal environment. The participants reported that they had from three to seven or more years’ experience with the En Route Automation Modernization (ERAM) system, or simply reported that they had been using ERAM since it was introduced to the field. All of the participants rated their air traffic control performance high, with a median response of 8 on a 10-point scale (1=low; 10 =high), and that they were highly motivated to participate in the simulation, with a median response of 10.

3.2 Facilities

We conducted the simulation at the RDHFL. The RDHFL includes two experiment rooms (ER2 and ER3) that emulate the en route air traffic control automation system and that contain en route consoles, have R-side and RA position displays, push-to-talk communication capabilities, and provide audio/video recording capabilities. Figure 1 presents a layout of the RDHFL simulation areas including ER2 and ER3.

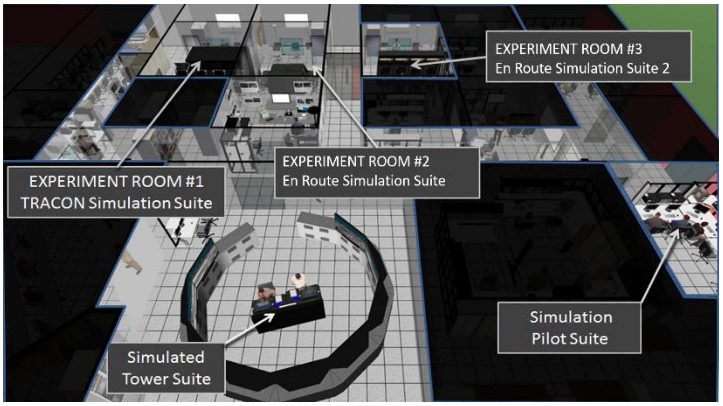


Figure 1. RDHFL Layout

3.3 Airspace

We used New York (ZNY) ARTCC sectors 27 and 10. We used this airspace because it had many of the characteristics desired for this simulation and had previously been adapted for use in simulations by researchers at the FAA Civil Aeromedical Institute (Crutchfield & Millan, 2017).

3.3.1 ZNY Sector 27

ZNY sector 27 is a low altitude sector that handles traffic into Philadelphia (KPHL) and satellite airports. The yellow highlighted section in Figure 2 depicts the location of sector 27 within ZNY, and Figure 3 depicts the altitude strata and boundaries of the airspace.

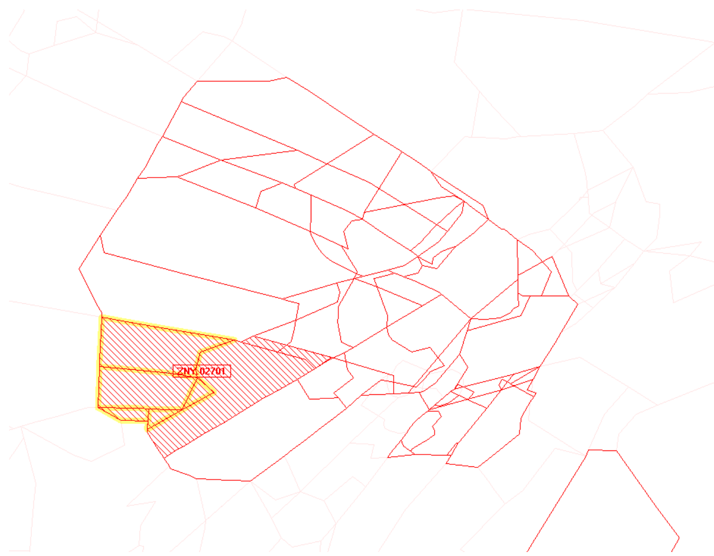


Figure 2. ZNY 27 low altitude sectors

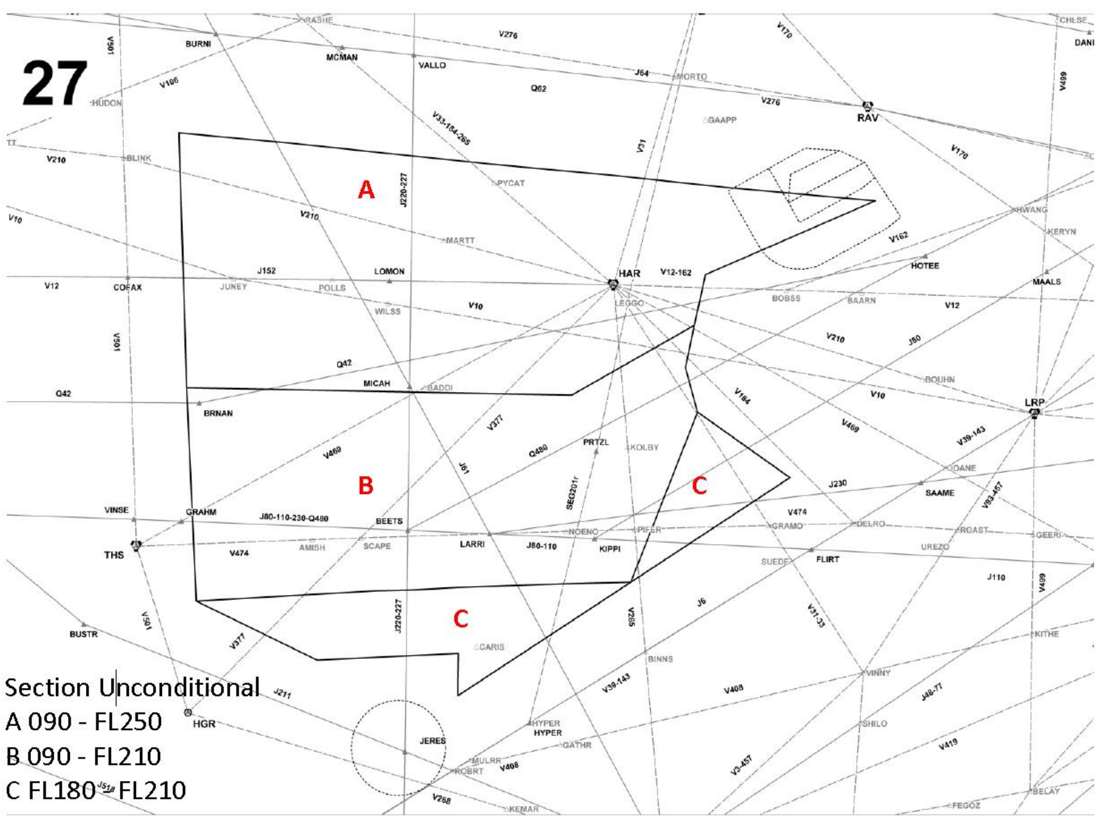


Figure 3. ZNY 27 altitude strata and boundaries

Figure 4 depicts the sectors surrounding ZNY 27 including the sector names, altitude strata, and voice communication frequencies.

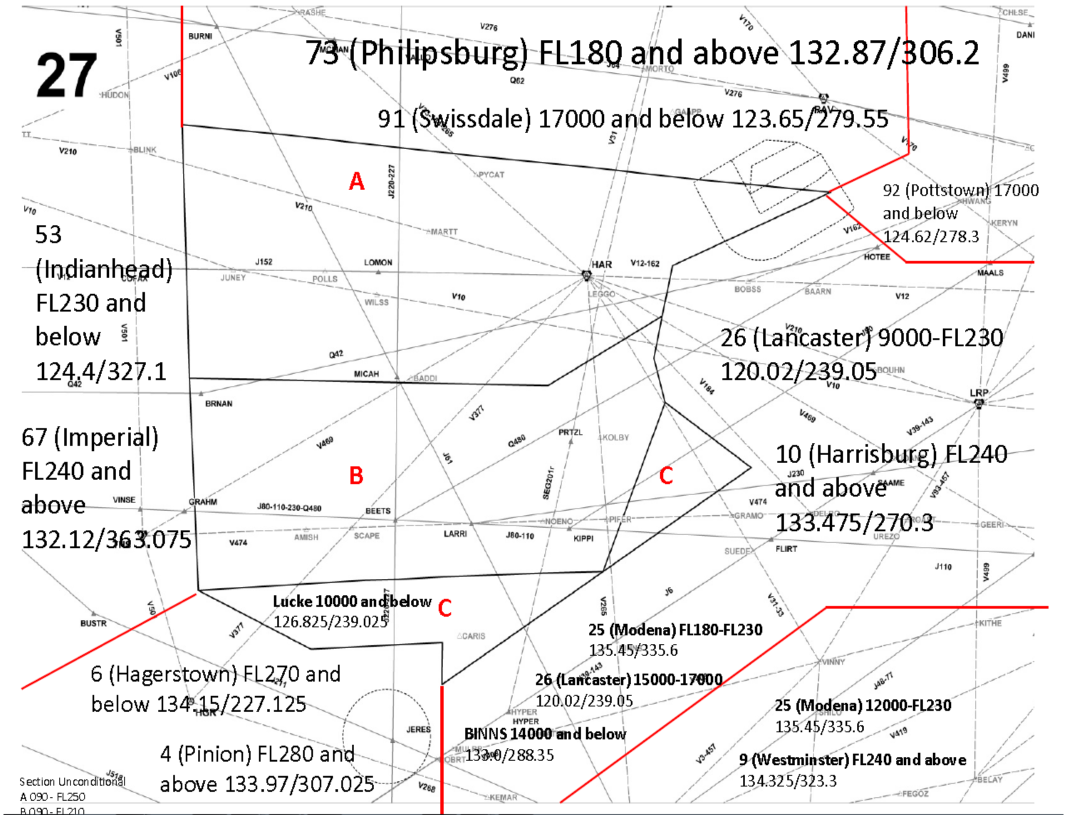


Figure 4. ZNY 27 surrounding sectors and voice communication frequencies

3.3.2 ZNY Sector 10

ZNY sector 10 is a high altitude sector with predominantly departure flows from Philadelphia (KPHL), the New York metropolitan area, and the Washington metropolitan area. The yellow highlighted section of the map in Figure 5 depicts the location of sector 10 within ZNY. Figure 6 depicts the altitude strata and boundaries of the airspace for ZNY 10. Figure 7 shows the sector names, altitude strata, and voice communication frequencies of the sectors surrounding ZNY 10.

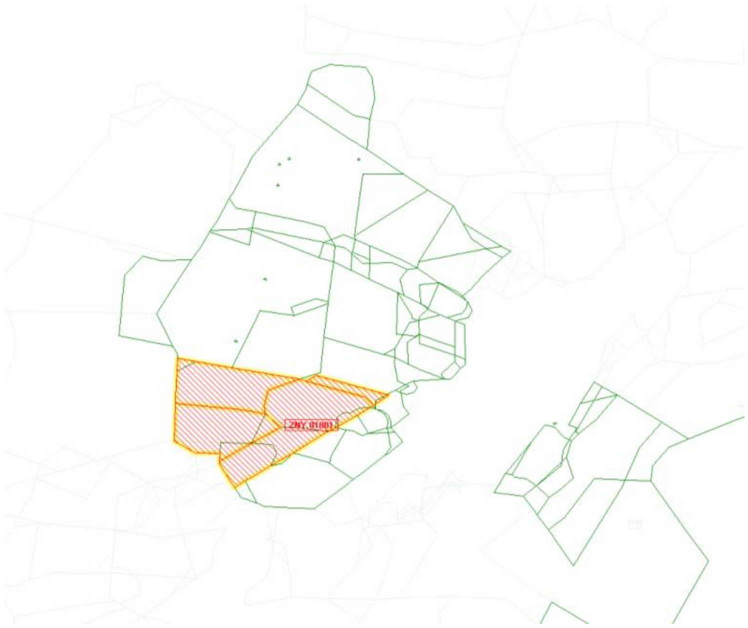


Figure 5. ZNY10 high altitude sectors

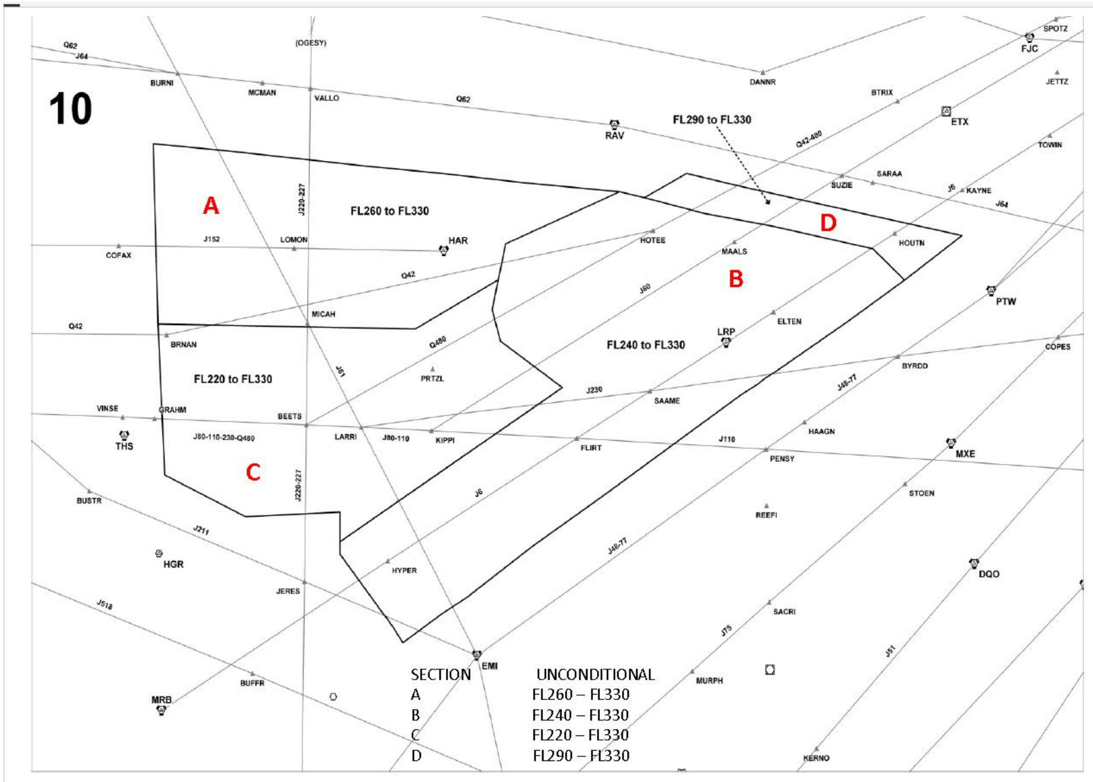


Figure 6. ZNY 10 altitude strata and boundaries

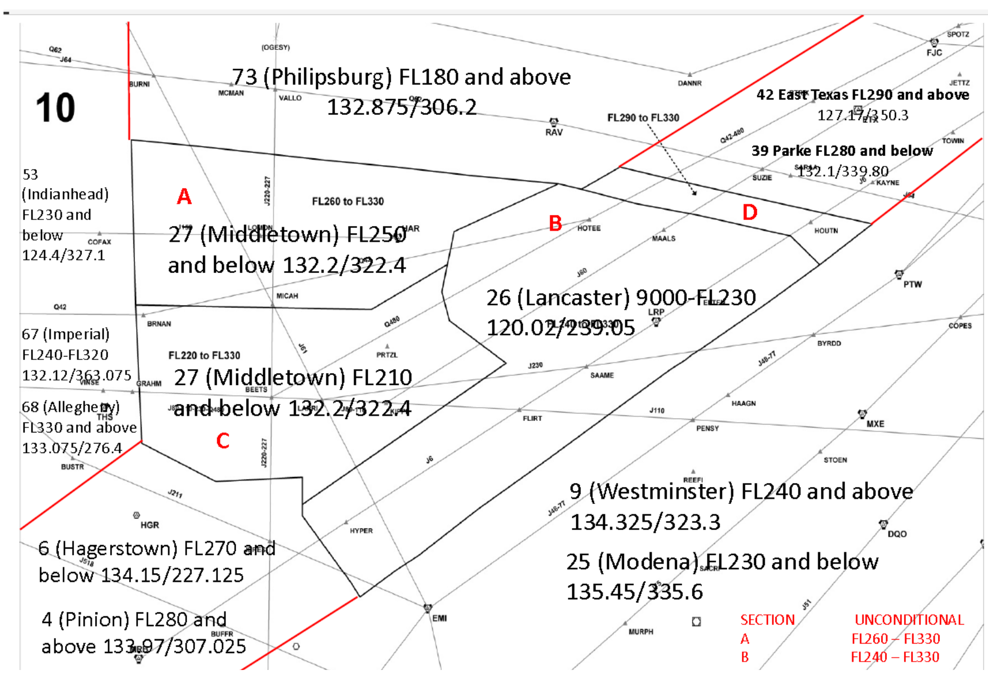


Figure 7. ZNY 10 surrounding sectors and voice communication frequencies

Section A and C of ZNY 10 were directly above ZNY 27 to allow transfer of traffic and coordination between the sectors.

3.4 Simulation Software and Laboratory Capabilities

We used several systems to create a high-fidelity en route simulation environment. The following sections describe each of these systems and the systems to which we connected. In the sections that follow, we provide an overview of each of the simulated capabilities and examples of display elements and menus.

3.4.1 Distributed Environment for Simulation, Rapid Engineering, and Experimentation (DESIREE)

To emulate an en route air traffic control automation platform, we used the Distributed Environment for Simulation, Rapid Engineering, and Experimentation (DESIREE), developed at the RDHFL. DESIREE is an air traffic control simulator that can simulate en route and terminal air traffic control automation. DESIREE receives input from the Target Generation Facility (TGF) to display radar targets and aircraft information on the controller displays, including radar tracks, data blocks, and sector maps. DESIREE allows controllers to perform functions as they would in an operational environment. DESIREE has data collection and storage capabilities to capture information pertaining to aircraft controller interactions with the system, such as making and accepting handoffs, data block entries, interacting with display elements, and so forth. DESIREE also connects to and coordinates all hardware and recording equipment. Figure 8 provides a depiction of the simulation environment used in this study, which includes the 43” monitor on the R-side, communication panel, and RA position capabilities.

Figure 8. Simulation Environment

3.4.2 Target Generation Facility and Simulation Pilots

The TGF software simulates realistic aircraft dynamics, provides aircraft position data to DESIREE, and includes simulation pilot workstations and software. The TGF software reads in airspace adaptation and air traffic flight plan data to generate the track data that it provides to DESIREE. The simulation pilot workstations were located in a separate room from ER2 and ER3 at the RDHFL. Either two or four simulation pilots were assigned to each sector depending upon the experiment condition.

3.4.3 Voice Communication and Switching System

The simulated Voice Communication and Switching System (VSCS) recorded communications between the controllers and the simulation pilots whenever they keyed and released the microphone. This allowed us to determine time of onset and duration of the push-to-talk (PTT) communications between participants and pilots.

3.4.4 Joint En Route Decision Support System

To implement Conflict Probe, DESIREE incorporates Conflict Detection capabilities provided by the Joint En Route Decision Support System Infrastructure (JEDI), a Linux-based User Request Evaluation Tool (URET) prototype maintained by the MITRE Corporation. JEDI detects and provides data about potential conflicts to DESIREE, which presents the information as Conflict Probe notifications. JEDI examines aircraft trajectories and probes for potential conflicts several minutes into the future. We used the default parameters: a 20-minute look-ahead time for aircraft-to-aircraft conflicts and 40-minute look-ahead time for aircraft-to-airspace conflicts.

This simulation included notifications on the RA position display as currently provided in the field, but also added the notifications and symbology anticipated for the R-side display. The R-side features included notifications of potential conflicts in line zero of the full data block (FDB) as well as in the Probe Alert List. The indicator provides two pieces of information: (a) the number of potential conflicts and (b) a color code for the type of conflict indicated—red or yellow for aircraft-to-aircraft conflicts and orange for aircraft-to-airspace conflicts. Figure 9 depicts the notifications as they appear on an FDB and in the Probe Alert List.

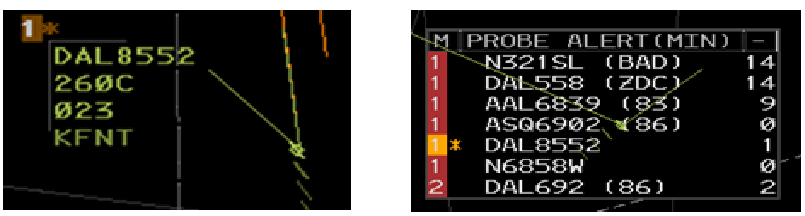


Figure 9. Conflict Probe, R-side:FDB (left);Probe Alert List (right)

Selecting a notification displayed the trajectories for the affected aircraft as shown in Figure 10.

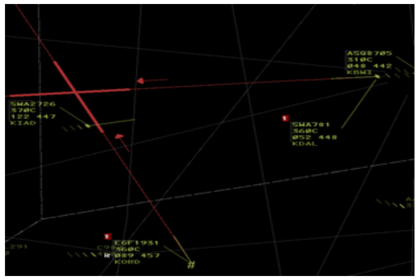


Figure 10. Trajectory probe graphic

3.4.5 Time-Based Flow Management

For the simulation, DESIREE interfaced with the operational TBFM software to generate speed advisories for aircraft to meet scheduled arrival times at fixes, metering points, and destinations. TBFM provides Delay Countdown Times (DCTs), Scheduled Times of Arrival (STA), and Ground-based Interval Management-Spacing (GIM-S) advisories for aircraft. The controller can accept or reject proposed advisories or let them time out. The speed advisory is provided via the coordination ("C") indicator in line 0 of the FDB, as depicted in Figure 11. When the 'SPEED ADVSRY' button is toggled to "on" in the main toolbar, speed advisory indicators appear on the Meter Reference Point (MRP) List, on the fourth line of the FDB, and in the Aircraft List on the RA position display, if applicable. We set the default for this option to "on" for this simulation.

When the controller selects the "C" indicator, the speed menu appears, displaying the proposed

Figure 11. TBFM GIM-S advisory

advisory (see Figure 12). The controller can select or reject the advisory, or allow it to time out.

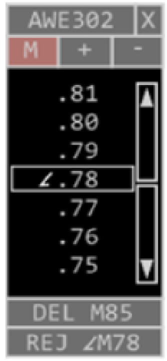


Figure 12. GIM-S speed menu

If TBFM determines that a speed advisory is not sufficient to enable the aircraft to meet the STA, it calculates a Path Stretch Advisory reroute to absorb the delay. A yellow hourglass appears in the FDB and MRP list next to the aircraft speed to indicate that a Path Stretch Advisory is available. The system probes for conflicts in calculating the proposed reroute. The proposed route and other information including the heading off the route, turn-back point, and rejoin point are provided on the controller display similar to that depicted in Figure 13.

Figure 13. Path Stretch Advisory

The controller can choose to accept or reject the advisory or create an alternative reroute option. Accepting the advisory automatically generates a flight plan amendment message for the path stretch maneuver. If the controller accepts the advisory, the hourglass becomes green.

In the simulation, DESIREE connected to the TBFM system through the En Route Data Distribution System (EDDS) software to exchange data to populate the TBFM software and to receive the TBFM data needed to integrate with the ERAM emulation platform. We used TBFM in Sector 27 to schedule arrival aircraft to Philadelphia airport (KPHL) and its satellite airports. ZNY 10 did not receive TBFM data.

In our simulation, the TBFM system provided only GIM-S speed advisories. If DESIREE determined that the speed advisory provided by TBFM could not be met, it sent routes to JEDI to evaluate route options to provide a Path Stretch Advisory. JEDI probed the routes to determine whether they were conflict free, and then DESIREE ranked them for use as Path Stretch Advisories. Due to limitations, JEDI was unable to account for Special Activity Airspace (SAA) in its evaluations for Path Stretch Advisory. This meant that Path Stretch Advisory routes could be displayed that entered the SAA. This outcome would not occur in the operational environment.

3.4.6 Data Communications

The DESIREE system simulated Data Communications (Data Comm) Full Services (DC-FS) display elements and functions as planned for the operational ERAM En Route Display System Management (EDSM). Data Comm allows controllers to send clearances to aircraft and receive requests from pilots using text, without using the voice frequency. The introduction of DC-FS will provide controllers with multiple new functions and symbols as shown in Table 1.

Table 1. Data Comm full services symbology

Indicator symbol name	Indicator	Examples in full data block (FDB)
Next Data Authority Session		APRØ1
Current Data Authority Session		SWA6Ø2
Transfer Of Communication (TOC) in progress (receiving)		SWAØ2Ø
Data Comm session with eligibility		APRØ1
Initial Contact in progress		USA9876
Initial Contact mismatch	39Ø	35Ø↑29Ø39Ø
On-frequency (auto)		3ØØC
On-frequency (manual)		3ØØC
Route uplink in progress	Ø15	Ø15 413
Route uplink -timeout	Ø34	Ø34V437
Route uplink-abnormally terminated	ØØ3	ØØ3 3ØØ
Altitude uplink in progress	26Ø	26Ø↑25Ø
Altitude uplink timeout	19Ø	19Ø↓2ØØ
Altitude uplink abnormally terminated	29Ø	29Ø↑28Ø
Emergency Pilot Initiated Downlink (PID) request received		UAL62
PID received		APRØ1
Generic uplink in progress		ROGER1
Generic uplink timeout		LINK1Ø
Abnormal generic uplink		COA999
Failed session with eligibility		SWAØ32
Held TOC single		USA9876
Held TOC Multiple		WALØ1
Held TOC abnormal		UAL66
TOC in progress transferring		USA9876
Abnormal TOC transferring		LINKØ6

We equipped 50% of the aircraft in the scenarios with Data Comm. However, at any time within a scenario, the percentage of Data Comm-equipped aircraft varied somewhat due to the dynamic nature of the traffic.

3.4.7 Airborne Reroute

DESIREE simulated the Airborne Reroute (ABRR) function that was introduced in ERAM EDSM to provide Traffic Flow Management (TFM) reroute information to the controller. The ABRR notification on the R-side is displayed as a "T" in the Range Data block (see Figure 14) that the controller can select via a trackball pick. Selecting the T brings up the probed route and the TFM Reroute Menu (see Figure 15), which indicates any protected segments of the route. In the simulation, the data provided in the route field (see Figure 16) for the R-side were also accessible from the Aircraft List (ACL) on the RA position display. However, this implementation is not yet available in the actual system.

Figure 14. ABRR in the range data block

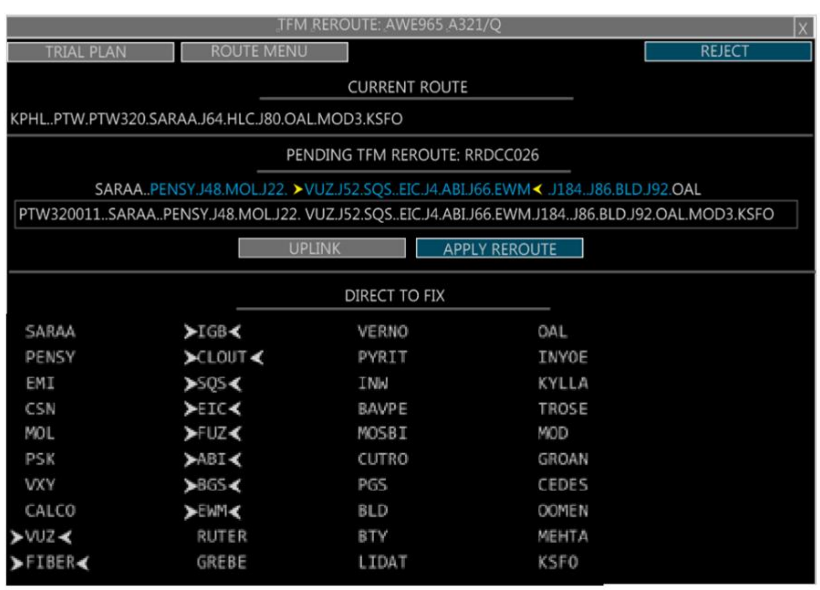


Figure 15. TFM reroute menu with protected segment (cyan)

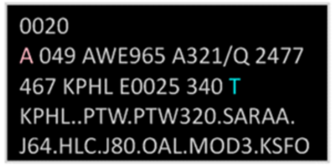


Figure 16. ABRR indicator in route field

The controller can trial plan the route to determine whether to accept, edit, or reject the reroute. The "T" is removed if the reroute is cancelled, rejected, or amended, or if the aircraft proceeds past a protected segment.

3.4.8 Controller-to-Controller Coordination

DESIREE simulated electronic Controller-to-Controller Coordination (CTCC) capabilities as planned for the field. CTCC allows a controller in one sector to make a request of a controller in another sector without using the landline, such as to request control to climb, descend, or turn an aircraft that is entering the sector.

The controller initiates electronic coordination by hovering over the speed portal on the FDB of the designated aircraft and selecting "Enter" on the trackball. A pop-up window displays available options. The available options depend on whether the selected aircraft is or is not under track control of the requesting controller, with fewer options available if the requesting controller does not have track control (see Figure 17).

Figure 17. CTCC options when controller has (left) or does not have (right) track control

After the controller selects and enters the desired option, a hollow green triangle appears on line 3 of the FDB, pointing away from the FDB to indicate that the request has been sent (see Figure 18). The controller receiving the request sees a hollow green triangle pointing toward the FDB on his display. Once requests are approved, the triangles are filled green. If the receiving controller chooses “unable” in response to a request, the triangle appears orange. If the controller chooses to counter the request (e.g., selects a different altitude than the one requested), the triangle appears yellow.

Figure 18. CTCC request sent (green triangle)

3.4.9 Additional En Route ATC Automation Functionality

For this simulation, we ensured that DESIREE implemented other capabilities that are currently available in the field and affect the way in which the controller interacts with the system. Most notable is the addition of a Voice Communications Indicator (VCI) to the radar display to help controllers monitor air/ground voice communications. The VCI is presented to the left of the altitude in the FDB. The frequency may be automatically marked, or the controller can manually mark an aircraft “on frequency” by selecting the blank location to the left of the call sign. Figure 19 presents the indicator as it appears for aircraft that have been automatically or manually marked “on frequency.”

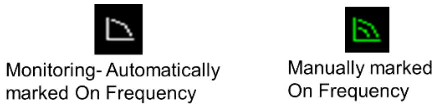


Figure 19. Automatically (left) and manually (right) marked frequency monitoring

We also included the capability for controllers to temporarily show information, such as beacon code and ground speed, in the aircraft representation that is otherwise only available from a flight plan readout or via information displayed in a time-shared field. Such on-demand information is available through press-and-hold buttons in the toolbar or as adapted on the keypad selection device. The beacon code and ground speed are displayed in Field E of the FDB. Other information available through a press-and-hold button is vertical rate (in hundreds of feet) which is displayed to the right of the 2nd line of the FDB.

3.5 Additional Equipment

We used other equipment in the simulation to elicit participant workload ratings and to monitor their eye movements and brain activity during the test scenarios.

3.5.1 Workload Assessment Keypad

We used the Workload Assessment Keypad (WAK) device to prompt participants to report their subjective workload during the scenarios. The WAK uses a 10-point scale that allow users to report their workload from very low (1) to very high (10). We provided the participants with the following definition of the ratings: Ratings of 1 and 2 indicate very low workload in which all tasks can be accomplished completely and easily. Ratings of 3, 4, and 5 indicate increasing levels of moderate workload in which all tasks can be accomplished but the chance for errors is increasing and there is less spare time available to accomplish all tasks. Ratings of 6, 7, and 8 indicate high workload in which no spare time is available, some unessential tasks go unfinished, and it is difficult to complete all essential tasks. Ratings of 9 and 10 indicate extremely high workload in which essential tasks go unfinished and separating aircraft is difficult. We encouraged the participants to use the full range of the scale. At the beginning of each day, we reviewed the rating scale definitions and reminded the participants to respond to the WAK prompts as much as possible. The WAK device is depicted in Figure 20 and is shown in Figure 8 above adjacent to the workstations.

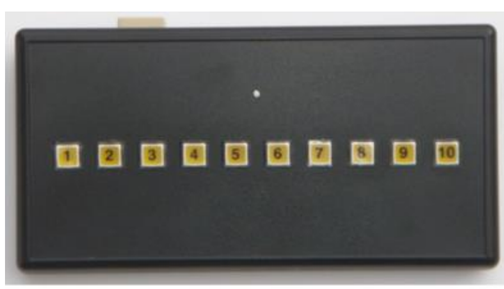


Figure 20. Workload Assessment Keypad

We configured the WAK to prompt participants for input every 2 minutes during the scenarios. At each prompt, the device emitted a tone and illuminated the 10 buttons. The participant had 20 seconds to respond. If the participant did not respond within 20 seconds, the WAK recorded a code for missing data. DESIREE recorded the responses (or missing data values) and the time at which the responses were made for later analysis.

3.5.2 Eye Tracker

We used the SmartEye Pro System to collect eye-tracking data from the participants when they worked the R-side positions for Sectors 10 and 27 in the test scenarios. The SmartEye Pro System has a large measurement volume that uses cameras to track head and eye movements in real time. The system does not require participants to wear any head-mounted gear, thus allowing them to move fairly freely. Figure 21 shows the SmartEye Pro System at an R-side position.

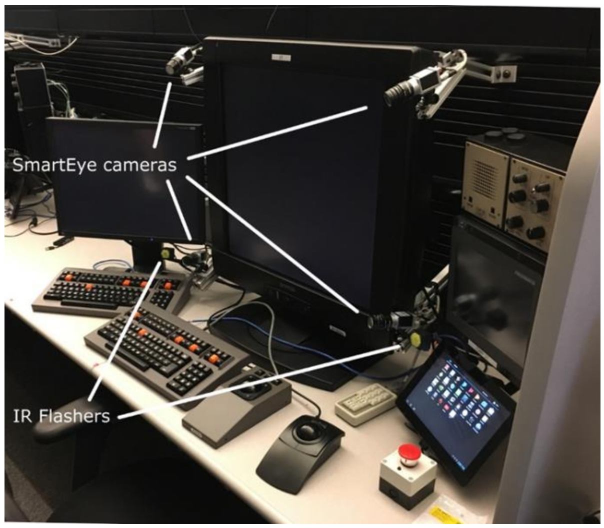


Figure 21. SmartEye Pro eyetracker at R-side posiiton

The SmartEye Pro system uses four cameras to capture the participant’s head as a 3-dimensional object at up to 120Hz. It determines the location of the eyes and the line of sight for each eye using near-infrared light. The intensity of the infrared illumination is about one thirtieth of the intensity encountered while walking outside on a sunny day and therefore causes little to no discomfort or health risk. We set up the system at the Sector 10 and Sector 27 R-side positions.

We configured SmartEye's 3D world model based on the surfaces of interest at the controller workstation. Figure 22 depicts the SmartEye 3D world view and the scene planes of interest that we defined: the R-side display, R-side keyboard, RA-position display, and RA-position keyboard. SmartEye measured the Point Of Gaze (POG) positions in pixels for the displays and in meters for the other scene planes.

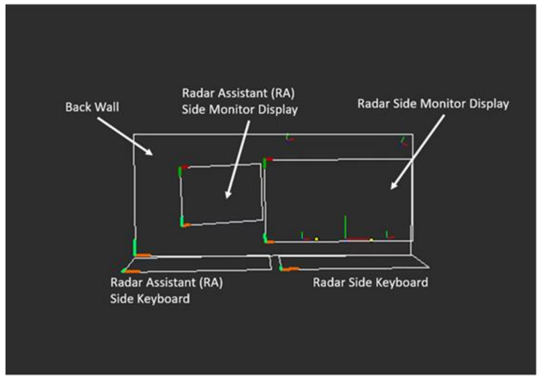


Figure 22. Scene planes for the R-side and RA positions

We obtained pupil diameter measurements from SmartEye Pro. This allowed us to correlate pupil diameter with POG on the aircraft data blocks. Researchers in previous research, including studies with air traffic controllers (Ahlstrom & Friedman-Berg, 2006), have found pupil diameter to be a measure of cognitive workload, with larger pupil diameters indicative of higher workload.

We wanted to measure whether pupil diameter increased as a function of FDB complexity. The data blocks in this simulation could potentially contain a large and varied amount of new information elements due to the number of new tools and capabilities. For example, depending on aircraft equipage and flight status, the participant could potentially see several new indicators for a single capability, such as Data Comm. For Data Comm, one indicator conveys information about equipage and availability, while other indicators are used to signify whether messages are being uplinked or downlinked, to signify abnormalities in message transmission, or to indicate that a sent message has timed out.

3.5.2.1 Full data block complexity

For this simulation, we defined data block complexity as the number of indicators that appeared on an FDB at the time an eye fixation occurred on that data block. The greater the number of indicators, the higher the complexity. We hypothesized that we would find that larger pupil diameters would correlate with higher complexity data blocks indicating a higher level of cognitive workload.

For the basic, and least complex FDB, we included the aircraft's position, position symbol, track symbol, leader line, call sign, assigned altitude, profile indicator, computer identification, and ground speed. Depending on aircraft status and equipage, additional elements appeared, including elements indicating Data Comm equipage and availability, uplink and downlink message indicators, voice and track control ownership indicators, transfer of communication status, heading and speed information (in the 4th line of the FDB), ADS-B status, safety indicators (conflict probe or conflict alert), CTCC status and coordination indications, point outs, Continuous Range Readout, delay countdown time, STA, flight event status information and non-RVSM status indicator. We provide examples of lower and higher complexity FDBs in Figure 23.

Figure 23. Examples of lower (left) and higher complexity (center and right) FDBs

In this simulation, the text information in the data blocks was always left justified and the portal fence—a white line that surrounds the text area—was not visible. All of the participants chose to use very low brightness levels in their preference settings for the portal fence, and therefore, it did not appear.

There are several other display elements that could potentially be presented in the FDB that we did not include in the simulation. These pertained to symbology for “abnormal” states (e.g., abnormal uplink) in which a red or orange box is presented around the primary symbol to indicate a problem. We did not include abnormal states in our simulation, so these indicators did not appear. Therefore, the FDBs in the operational environment could be even more complex than those in our simulation.

3.5.3 Electroencephalogram

We used EEG recording equipment to obtain measures of participant brain activity during the test scenarios when they worked the R-side positions at Sector 10 and Sector 27. We recorded the EEG data with a 32-channel Brain Vision ActiCHamp active electrode system from Brain Products Inc. We collected data from 30 scalp positions according to the international 10-20 positioning system (Chatrian, Lettich, & Nelson, 1988). To further capture the effect of oculomotor activity on the EEG signals, we placed two electrodes around the eyes laterally to record eye movements and blinks.

Figure 24 depicts an image of the EEG recording cap with electrodes. This EEG system uses a conductive gel to ensure contact between the scalp and the electrodes. The researchers provided

Figure 24. An EEG cap on the head

shampoo and towels to allow the participants to remove any remaining gel from their hair at the end of the test day before they left the lab building.

We monitored and recorded the EEG data using BrainVision PyCorder software and one signal amplifier for each of the two participants. Each unit was equipped with 32 Ag/AgCl impedance-optimized electrodes to sample and record electrical activity at 1000 Hz with a battery-powered 24-bit amplifier. We used a software bandpass filter set between 0.01-100Hz during recording. We injected each electrode with SuperVisc electro-conductive gel, manufactured by EASYCAP GmbH, which enabled us to collect data without any special scalp preparation. We placed the electrodes in EASYCAP modular EEG recording caps. The electrodes covered the frontal, central, parietal, and occipital regions of the participants' heads. We also placed ground electrodes on the midline of the skull over the pre-frontal lobe (Fpz point) and a reference electrode over the frontal lobe (Fz point). Figure 25 shows the approximate electrode locations.

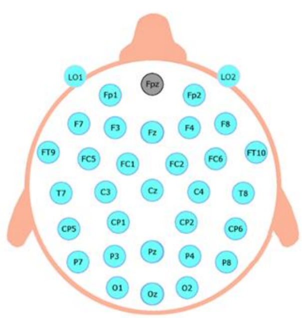


Figure 25. Electrode locations

The researchers applied the caps and electrodes before the first scenario on each test day and the participants wore them throughout the day. We adjusted the electrodes to get measured impedances below 10 KOhm. We ran impedance checks between sessions to monitor for loose or disconnected electrodes. The participants wore the audio headsets over the caps to communicate with the simulation pilots.

We connected each amplifier to a Microsoft Windows 10, 64-bit computer. DESIREE sent a trigger code to each amplifier, via a parallel port, to demarcate the beginning and end of a recording session. Additionally, DESIREE sent a trigger code every 60 seconds to ensure connectivity and synchronicity. We used an ActiMove wireless transmitter to send the EEG signal wirelessly from the participant to the recording computer.

We collected 128 recording files, eight from each of the 16 participants for the experimental scenarios in our analysis. After the first two scenarios of each testing day, the participants switched R-side positions and EEG wireless transmitters and amplifiers, but still wore the same caps and electrodes. Two of the four testing days also had participants positioned at the RA position next to the recorded participant. The participants at the RA position did not wear the EEG apparatus.

3.6 Scenarios

We based our traffic scenarios on the traffic samples obtained from ZNY Sectors 10 and 27 created by (Crutchfield & Millan, 2017) for use in simulations. During initial familiarization at the lab, the participants worked very low traffic scenarios so that they could get accustomed to the laboratory environment, the airspace, and the procedures, and become familiar with the appearance of the new tools and capabilities on the display. The participants completed a total of 8 familiarization scenarios before beginning the training scenarios.

For the training scenarios and half of the test scenarios, we included traffic levels that were close to the Monitor Alert Parameter (MAP) values for ZNY Sector 10 (15 aircraft) and 27 (12 aircraft). MAP values represent the maximum number of aircraft expected to be managed within a sector at a time. For the other half of the test scenarios, we included higher traffic levels (about 130% of the MAP values) for each sector to enable us to compare performance between different traffic levels. We designated these test scenarios as either having a moderate (M) or busy (B) traffic level in our simulation.

We equipped 50% of the aircraft with Data Comm, so we expected that voice communications workload would be somewhat reduced even though the traffic levels were often higher than what would typically occur in the field. Each of the scenarios ran for 40 minutes in duration. Every scenario displayed a “blank” screen for the first 3 minutes until the system loaded all of the traffic for the start of the scenario and made it available simultaneously on the sector display.

3.7 Materials

3.7.1 Informed Consent Statement

Each participant read and signed an Informed Consent Statement before the simulation (see appendix A). The Informed Consent Statement described the purpose of the study and the rights and responsibilities of the participants, including that participation was voluntary and that they could end their participation at any time without penalty. The document also informed the participants that the researchers would keep their data anonymous and confidential; only code numbers would be associated with the data, not names or identities, and that only the research team members would access the audio/video recordings of the simulation to review events for later analyses.

3.7.2 Background Questionnaire

Each participant completed the Background Questionnaire before the simulation (see appendix B). The Background Questionnaire asked participants to indicate their age, years of experience in air traffic control, and the amount of experience they had with ERAM.

3.7.3 Post-Scenario Questionnaire

At the conclusion of each test scenario, the participants completed a Post-Scenario Questionnaire (PSQ; see appendix C). The PSQ asked participants to rate their performance, situation awareness, and workload. The PSQ also asked questions about the understandability of the display elements and the extent to which the information presented during the scenario affected their ability (positively or negatively) to manage the traffic, maintain situation awareness, and resolve potential conflicts.

3.7.4 Exit Questionnaire

At the conclusion of the simulation, the participants completed the Exit Questionnaire (see appendix D). The Exit Questionnaire items asked the participants about the realism of the simulation, the efficacy of training, and the utility of the new tools and capabilities.

3.7.5 Debrief Session

After the participants completed the simulation, the researchers met with them to elicit additional feedback on the simulation and the new tools and concepts. We gathered and categorized the comments to provide additional rationale for the questionnaire responses.

3.8 Procedure

The participants arrived at the RDHFL in groups of four and completed the simulation over a 2-week period. They traveled in on Monday of the first week and traveled out on Friday of the following week. Two of the participants in each group had 5 or fewer years of experience as CPCs and the other two had 15 or more years of experience as CPCs.

When the participants arrived at the RDHFL, one of the researchers gave an introductory briefing that described the general purpose of the simulation, an overview of the new air traffic tools and capabilities, and discussed the schedule and logistics. The researchers provided a description of the data collection and recording equipment, including the WAK, eye tracking system, and EEG equipment, and informed the participants that all scenarios would be video and audio recorded for later review, as needed by the research team.

After the introductory briefing, the participants had the opportunity to ask questions, and then they read and signed the Informed Consent Statement. One of the researchers and a witness also signed the document. Next, the participants completed the Background Questionnaire. Then, the ATC SMEs instructed the participants on the airspace, Standard Operating Procedures (SOPs), and Letters of Agreement (LOAs) that would be used. The researchers and SMEs addressed any initial questions and then proceeded to the lab to begin familiarization on the laboratory equipment, airspace, and procedures. Each participant took turns as the R-side for Sector 10 and the R-side for Sector 27 during the familiarization scenarios. All of the scenarios, including those used for familiarization, ran for 40 minutes. We took 15–20 minute breaks after each scenario and a 1-hour lunch break in the middle of the day during both weeks of the simulation. We began the simulation at 8:00 AM and completed training or testing by 4:30 PM each day.

Following initial familiarization, the SMEs introduced the new tools and concepts to the participants in modules over the first two days at the laboratory so they could learn to use them over time. For each training module, the SMEs initially provided a “classroom” briefing that included an overview of the tool via a slideshow with images and descriptions. The participants then went to the laboratory to gain experience working with the new tool. We provided instructions for the tools in the following order: CTCC, Conflict Probe on the R-side, Data Comm, TBFM, and ABRR.

In the laboratory, the researchers and SMEs provided additional guidance on the use of the tools and answered any questions. The participants worked as R-sides alone for the first 2 days of familiarization and training on each of the modules. They continued to work with all of the tools and capabilities in all of the remaining training sessions on days 3 and 4, with continued guidance and feedback from the researchers and SMEs as needed. On days 3 and 4, the participants also worked as R-side/RA-side teams to become familiar working in that configuration.

Each of the participants took turns working as the R-side for Sector 10 and for Sector 27 during the training modules on days 1 and 2. We counterbalanced the order in which they controlled each sector for each module. We conducted the training sessions for each module in both ER2 and ER3 so that all four participants could work at the R-side positions simultaneously. Two of the participants worked Sector 10 and Sector 27 in ER3 and two of the participants worked Sector 10 and 27 in ER2. We ran the same scenarios in ER2 and ER3 but the traffic was independent, and the participants and simulation pilots assigned to one room could not communicate with the participants and simulation pilots assigned to the other room.

After completing all of the modules, the participants continued working another 12 training scenarios on days 3 and 4. When the participants worked as R-sides alone, we used both ER2 and ER3, and we used ER3 for scenarios in which the participants worked in R-side/RA-side teams. We provide an example of the schedule depicting a sequence of familiarization and training scenarios for the first week of the simulation inappendix E.

We encouraged the participants to use the new tools and capabilities as much as possible and to ask questions of the SMEs and the researchers during the familiarization and training week. We asked the participants to respond to the WAK ratings during the last 2 days of training so that they could become accustomed to using it and build it into their work routine. We reviewed the rating scale definitions each morning. We included eye tracking measures during the last 2 days of training so that the participants could become familiar with the calibration procedure prior to the start of the scenario and so that we could determine the position of the participant at the workstation that would provide optimal system responsiveness.

The participants spent the second week of the simulation (4 days) completing the test scenarios. During the testing week, the researchers and SMEs did not provide assistance or direction on the use of the tools other than to answer any remaining questions prior to the start of the first test scenario. We ran four test scenarios per day for our core analyses. We ran one additional exploratory scenario, if time allowed, at the end of the day in which we had a participant from the Low Experience group and a participant from the High Experience group work together as a team. Test days required more time for preparation before the start of a scenario than practice days. On test days, we applied the EEG apparatus to the R-side participants when they arrived at the lab. This procedure took approximately one hour. We also calibrated the eye tracker prior to each scenario which took several minutes.

We counterbalanced the order of the test scenarios (moderate [M] or busy [B]) and the team configurations (R-side alone or R-side with RA) for each of the groups to minimize the possible effects of learning or fatigue. However, we did so with the requirement that the participants wearing the EEG equipment would wear it for a full test day due to the time-consuming process of applying, removing, cleaning, and (if needed) re-applying the device. Therefore, the participants worked at the R-side position for both Sector 10 and 27, with or without an RA (depending on test order) when wearing the EEG equipment, and they did not work at either of the RA positions on those days.

We conducted the test scenarios in ER3, the laboratory in which the EEG and eye tracking equipment were located. When the participants worked as R-sides alone, those designated to complete the test scenarios did so in ER3 while the other participants completed the same scenarios as R-sides in ER2 to equate the total time spent controlling traffic. We did not include the data from the ER2 scenarios as part of our analyses since ER2 as not equipped with Eye-trackers and EEG recording equipment. When the participants worked as R-side/RA teams, we ran only in ER3 in order to record EEG and eye tracking data from the participants working the R-side positions. Each participant completed a total of 16 test scenarios, eight of which were in ER3 and used for analysis. An example of the test schedule for the first two days of testing for one of the groups is provided in appendix F.

At the conclusion of each 40-minute test scenario, the participants completed the PSQ and then took a 20-minute break before the next scenario began. The participants completed the Exit Questionnaire after completing all of the test scenarios. The researchers held a debriefing session at the end of the simulation to allow the participants to discuss their experiences and provide additional comments about the new tools and capabilities.

4 Experimental Design and Analysis

Our objective was to understand how co-locating new tools and capabilities on the existing en route ATC automation system affects air traffic controller performance and workload. This simulation was a first step in obtaining baseline information toward that objective. We wanted to determine whether the participants made use of the new tools and capabilities and if they found them helpful in supporting their tasks or found them confusing, distracting, or difficult to work with. We wanted to determine whether controllers with different experience levels would make use of the tools differently or react to them differently. We also wanted to evaluate whether different traffic levels and different staffing configurations affected these measures.

We coded all data by participant number, not by name or identity, to ensure confidentiality. Our experiment used a 2 x 2 x 2 mixed design with participant Experience Level (Low and High) as a between-subjects factor and Scenario (Moderate or Busy) and Team Configuration (R-side alone and R-side with RA) as within-subjects factors. We report only significant effects, those for which we found *p* values less than .05. Our primary analysis tool was a 2 x 2 x 2 mixed model Analysis of Variance (ANOVA). We report the *F* and *p* values for significant effects, as well as partial eta squared (η^2) values as a measure of effect size. Partial (η^2) reflects the amount of variance of the dependent variable explained by the independent variable (e.g., scenario) and that partials out the effect of the other independent variables found not to have an effect (e.g., team configuration). The higher the value, the greater the effect size, with a value of .01 indicating a small effect, .09 a medium effect, and .25 or greater a large effect.

Some of the data sets involved measures that were not amenable to analysis by ANOVA because the data are often not normally distributed. These data sets included the questionnaire ratings and some of the data sets that included frequency counts (and could include values of zero). Therefore, we conducted non-parametric analyses for these data sets that we describe in the relevant sections below.

We analyzed the data separately for Sectors 10 and 27 because they already differed in their airspace size and configuration, the type of traffic handled, and the procedures and tools used. We analyzed the data in each scenario beginning 3 minutes into the scenario because that was the time at which traffic in the sector initially appeared on the controller display. We completed the analyses at 40 minutes, the end of the scenario.

We hypothesized that the participants would use the new tools and capabilities differently as a function of their perceived usefulness and as a function of perceived workload. The participants may use the tools infrequently if their workload was high. Alternatively, they may use the tools more frequently when their workload was high if the tools helped them better manage the traffic. We evaluated workload in several ways: (1) via the WAK ratings, (2) via the questionnaire ratings, (3) via pupil diameter data, and (4) via EEG data. We also looked for evidence of workload in the air traffic data and the efficiency and safety with which the participants handled the aircraft.

We used the EEG data in this study to determine whether we could identify physiologically-based evidence of high cognitive workload and possibly mental saturation levels. Due to the number of new tools and capabilities and the increased amount of information that could be presented in the aircraft FDB, the data blocks could vary in their complexity levels. We defined complexity by the number of elements available in the FDB at the time the participant fixated that FDB. The data include static information (e.g., aircraft call sign and type) and dynamic information (e.g. altitude and heading) as well as the additional elements added for the new tools and capabilities. We correlated the EEG and pupil diameter data to determine whether we would see EEG activity level increase as data block complexity increased and then plateau at a certain point to suggest that maximum capacity, or saturation level, had been reached. We discuss our approach to analyzing the EEG data in the results section below.

5 Results

We summarize the results to address the efficiency and safety with which the aircraft were managed, the perceived workload encountered, the extent of the use of the new tools and capabilities, and the metrics obtained for EEG and eye tracking as a function of the simulation test conditions. The air traffic measures included the number of aircraft managed in each sector, how near the sector boundary handoffs were made or received, the number of push-to-talk (PTT) transmissions, and any deviations from airspace procedures or losses of separation (LOS) that occurred. Subjective measures included ratings of workload and responses to questionnaire items. The EEG and eye tracking data also served as indicators of cognitive workload. Additional EEG analyses allowed us to explore the possible regions of the brain most involved in the processing of the FDBs.

5.1 Air Traffic Measures

We evaluated several air traffic measures to determine whether the experimental conditions affected the efficiency and safety with which the participants managed the aircraft through the sectors. As previously noted, we conducted the analyses for Sector 10 and Sector 27 separately because the characteristics of the sectors and the tools available differed between them. TBFM was only available in Sector 27 because only that sector handled arrival traffic metered to airports. The other tools were available in both sectors.

We evaluated the participants' ability to manage traffic through the sector to determine the effects of Experience Level (Low or High), Scenario (M or B), and Team Configuration (R-side alone or R-side with an RA). We evaluated the average number of aircraft handled, the time and distance of handoffs from the sector boundary for aircraft entering and exiting the sector, the time and distance aircraft were in the sector, the number of push-to-talk communications, any losses of separation (LOS) observed, and the number of airspace procedure deviations found.

5.1.1 Number of Aircraft

We evaluated the average number of aircraft managed in the sectors throughout the scenario to obtain a measure of how much busier the B scenario was than the M scenario. We obtained the overall average number of aircraft handled for each scenario by calculating the average number across each of the 2-minute intervals.

As expected, for Sector 10, we found a significant effect of scenario with more aircraft in the sector on average during a 2-minute interval in the B scenario (M = 11.69, SD = 0.528) than in the M scenario (M = 10.72, SD = 0.532), $F(1,14) = 35.49$, $p < .001$, partial $\eta^2 = .72$. Likewise, as expected, we found a significant effect of scenario for Sector 27 with more aircraft in the sector on average during a 2-minute interval in the B scenario (M =12.05, SD = 0.616) than in the M scenario (M = 11.06, SD = 0.447), $F(1,14) = 36.37$, $p < .001$; partial $\eta^2 = .72$. Neither the effect of Experience Level nor Team Configuration were significant.

5.1.2 Handoffs

Overall, when working in a team configuration, the participants accepted handoffs into their sector as R-side controllers most of the time (96%), whereas handoffs out of the sector were more evenly distributed between the R-side and RA participants (approximately 60% and 40%, respectively).

We evaluated when the R-side participants accepted handoffs for aircraft entering the sector and when they made handoffs for aircraft exiting the sector. We also evaluated the distance of the aircraft from the sector boundaries. Handoffs accepted closer to the sector boundary or with less time from the sector boundary may indicate that the participant is busy or experiencing a higher level of workload. For outbound aircraft, handoffs made earlier and further from the sector boundary may indicate that the participant is looking to reduce workload by moving aircraft out of the sector as soon as possible.

We evaluated the time and distance at which handoffs were taken or made by the R-side participant from the sector boundary for aircraft entering and exiting the sector separately for Sector 10 and Sector 27.

5.1.2.1 Sector 10 handoffs

For handoffs accepted into Sector 10, we found a significant interaction of Scenario x Team Configuration for the distance at which handoffs were taken by the R-side participant for aircraft entering the sector, $F(1,14) = 6.45$, $p = .023$, partial $\eta^2 = .315$. Figure 26 depicts the mean distance of the aircraft from the Sector 10 boundary for handoffs accepted. When working alone, the participants accepted handoffs further from the sector boundary in the M scenario than in the B scenario ($p < .001$). This difference was not as great when they worked with an RA ($p < .05$).

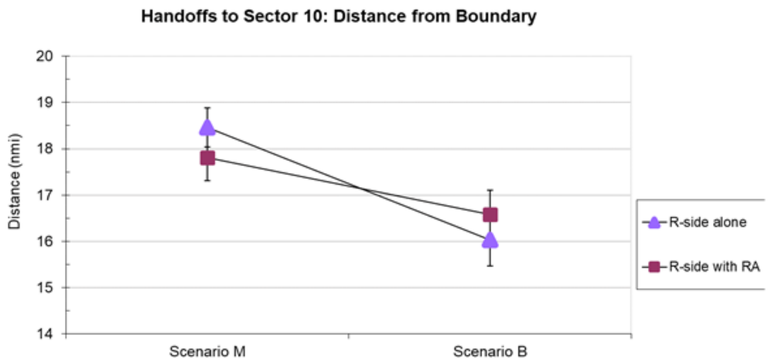


Figure 26. Handoff acceptance: distance from Sector 10 boundary

We also found a significant interaction of Scenario x Team Configuration for the time at which handoffs were taken for aircraft entering the sector, $F(1,14) = 6.22$, $p = .026$, partial $\eta^2 = .308$. These results were similar to the handoff acceptance distance results above and are shown in Figure 27. When working alone, the participants accepted handoffs with more time (s) from the sector boundary in the M scenario than in the B scenario ($p < .001$). The difference was not as great when they worked with an RA ($p < .05$).

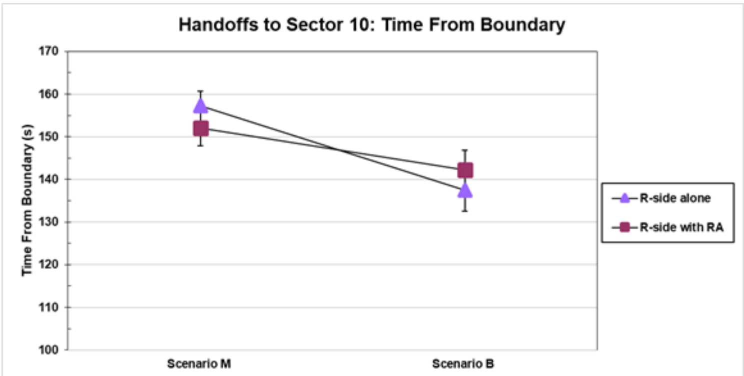


Figure 27. Handoff acceptance: time from Sector 10 boundary

For handoffs made from Sector 10, the participants handed off the aircraft closer to the boundary when working the M scenario (M = 14.28, SD = 1.25) than they did when working the B scenario (M = 15.70, SD = 1.54), $F(1,14) = 24.76$, $p < .001$, partial $\eta^2 = .639$.

We also found a significant interaction of Scenario x Team Configuration for the time at which aircraft were handed off from Sector 10, $F(1,14) = 5.092$, $p = .041$, partial $\eta^2 = .267$. Figure 28 depicts the mean time of handoffs from Sector 10. The participant handed off the aircraft latest—with the least time (s) from the sector boundary—when they worked the M scenario alone ($p < .001$) compared to when they worked the M scenario with an RA or when they worked the B scenario alone or with an RA ($p < .001$).

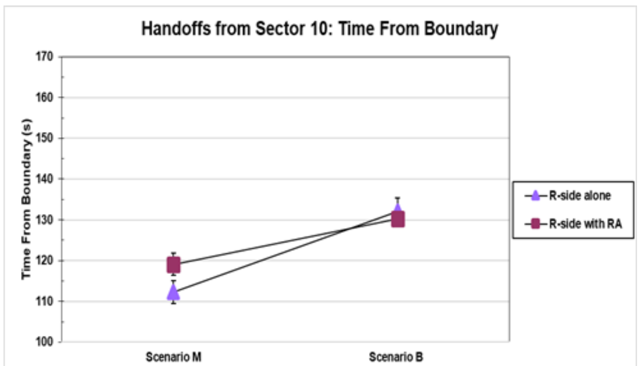


Figure 28. Handoffs from Sector 10: time from boundary

The results for Sector 10 suggest that participants were willing to accept aircraft into the sector sooner when traffic levels were moderate and more likely to handoff earlier when traffic levels were higher especially when they worked as R-side alone.

5.1.2.2 Sector 27 handoffs

The participants took handoffs into Sector 27 when aircraft were further from the sector boundary (nmi) when working the M scenario (M = 12.23, SD = 2.06) than when working the B scenario (M = 9.98, SD = 2.73), $F(1,14) = 72.696$, $p < 0.001$, partial $\eta^2 = .839$. The participants took handoffs further from the sector boundary when working as an R-side with an RA (M = 11.65, SD = 2.35) than they did when working alone (M = 10.55, SD = 2.54, $F(1,14) = 11.376$, p =.005, partial $\eta^2 = .447$.

We found similar results for the time (s) at which aircraft handoffs were accepted into Sector 27. The participants accepted handoffs into the sector later—with less time until the aircraft reached the boundary (s)—when working the B scenario (M = 99.67, SD = 26.58) than they did when working the M scenario (M = 120.93, SD = 18.99), $F(1,14) = 50.291$, $p < .0001$, partial $\eta^2 = .782$. The participants also accepted handoffs into the sector later when working as an R-side alone (*M* = 104.94, *SD* = 24.43) than when working as an R-side with an RA (*M* = 115.65, *SD* = 21.80), $F(1,14) = 11.947$, $p = 0.004$, partial $\eta^2 = .460$.

For handoffs made from Sector 27, we found a significant main effect of Scenario $F(1,14) = 5.629$, $p = 0.033$, partial $\eta^2 = .287$, on the distance (nmi) from the boundary at which handoffs were initiated. The participants handed off the aircraft further from the sector boundary when working the B scenario (*M* = 8.49, *SD* = 0.75) than when working the M scenario (*M* = 8.19, *SD* = 1.02). We did not find any significant results for handoff time from the boundary.

The results for Sector 27 suggest that participants were willing to accept aircraft into the sector sooner when traffic levels were moderate and more likely to handoff earlier when traffic levels were higher. They were also more willing to accept aircraft into the sector sooner and handoff aircraft later when they worked in R-side/RA teams.

5.1.3 Aircraft Time and Distance in Sector

We evaluated the total time and distance that aircraft were in the sector under the participants' control (see Table 2). For Sector 10, the aircraft traveled a longer distance (nmi) under the participants' control in the M scenario than in the B scenario, $F(1,14) = 14.87$, $p = .0017$, partial $\eta^2 = .515$. Likewise, the aircraft traveled for a longer duration under the participants' control in the M scenario than in the B scenario, $F(1,14) = 8.93$, $p = 0098$; partial $\eta^2 = .389$.

Table 2. Aircraft time and distance in sector

Sector	**Scenario**	**Mean Distance (SD)**	**Mean Time (SD)**
ZNY10	M	54.849 nmi (1.99)	7.621 min (0.29)
	B	52.753 nmi (1.98)	7.407 min (0.26)
ZNY27	M	42.784 nmi (1.35)	7.343 min (0.20)
	B	39.967 nmi (1.55)	6.748 min (0.26)

For sector 27, the aircraft again traveled a longer distance under the participants' control in the M scenario than in the B scenario, $F(1,14) = 69.26$, $p < .001$, partial $\eta^2 = .831$. We also found a significant interaction of Experience Level x Scenario on aircraft time (min) in the airspace, $F(1,14) = 6.39$, $p = .024$; partial $\eta^2 = .313$ as shown in Figure 29. The participants in the Low Experience group handled aircraft for a longer duration when working the M scenario (*M* = 7.48, *SD* = 0.07) than the participants in the High Experience group (*M* = 7.21, *SD* = 0.10), $p < .005$, but the groups did not differ significantly from one another when working the B scenario ($p > .05$).

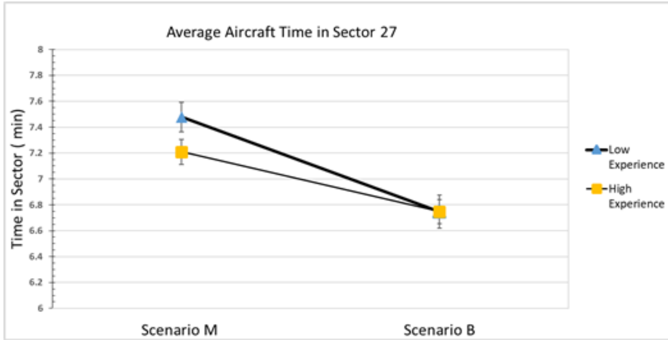


Figure 29. Mean duration of aircraft in Sector 27

Overall, the participants were more likely to handle the aircraft for a longer duration when traffic levels were moderate.

5.1.4 Push-to-Talk Transmissions

We analyzed the number and duration of controller-to-pilot and pilot-to-controller Push-To-Talk (PTT) transmissions in each scenario. We eliminated transmissions that were less than 150 msec in duration because it would not have been possible to issue a meaningful communication within that time. We equipped one half of the aircraft in all of the scenarios with Data Comm, so the participants could use voice or Data Comm to communicate with these aircraft, although we expected that the participants would use voice transmissions to issue commands that required more timely compliance.

For Sector 10, we found a significant interaction of Scenario x Experience Level, $F(1,14) = 4.652$, $p = .049$, partial $\eta^2 = .249$, on the number of controller-to-pilot transmissions as shown in Figure 30. The number of transmissions was higher in Scenario B than Scenario M, $F(1,14) = 418.49$, $p < .001$, partial $\eta^2 = .968$, but the High Experience group made more transmissions when working the M scenario than did the Low Experience group ($p < .05$). Both groups made the same number of transmissions in Scenario B.

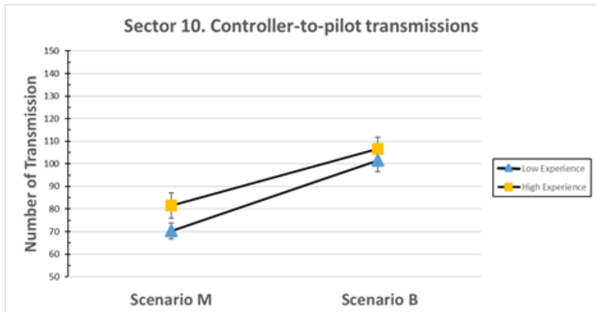


Figure 30. Sector 10: Average number of controller-to-pilot transmissions

Although the differences were small, we found a significant interaction of Scenario x Team Configuration on the duration of controller-to-pilot transmissions in Sector 10, $F(1,14) = 5.196$, $p = .029$, partial $\eta^2 = .297$. When the participants worked as a team, they made longer transmissions in the B scenario than the M scenario, but the duration of their transmissions was the same in the B and M scenarios when working alone (see Figure 31).

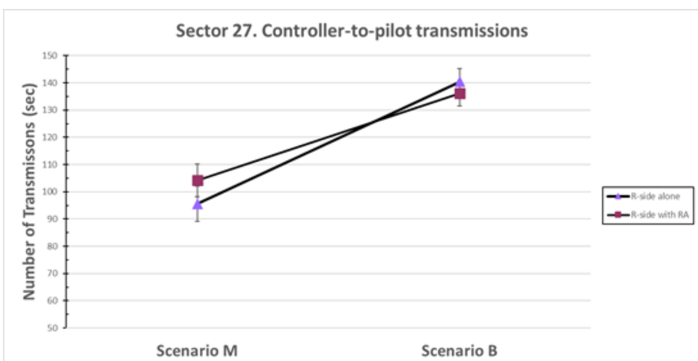


Figure 31. Sector 27: Average number of controller-to-pilot transmissions

For Sector 10 pilot-to-controller transmissions, we found more transmissions made in the B scenario ($M = 124.69$, $SD = 12.72$) than the M scenario ($M = 88.88$, $SD = 10.68$), $F(1,14) = 464.59$, $p < .001$, partial $\eta^2 = .971$ (a result we expected given the greater number of aircraft in the B scenario). We did not find any significant effects of the test conditions on pilot-to-controller transmission durations which had an overall average duration of 3.69 s ($SD = 0.36$).

For Sector 27, we found a significant interaction of Scenario x Team Configuration, $F(1,14) = 5.00$, $p = .042$, partial $\eta^2 = .263$, on the number of controller-to-pilot transmissions as shown in

Figure 32. The participants made more transmissions when working the B scenario than the M scenario, $F(1,14) = 111.928$, $p < .001$, partial $\eta^2 = .889$, but the difference was greater when they worked as R-sides alone than when they worked with an RA ($p < .05$).

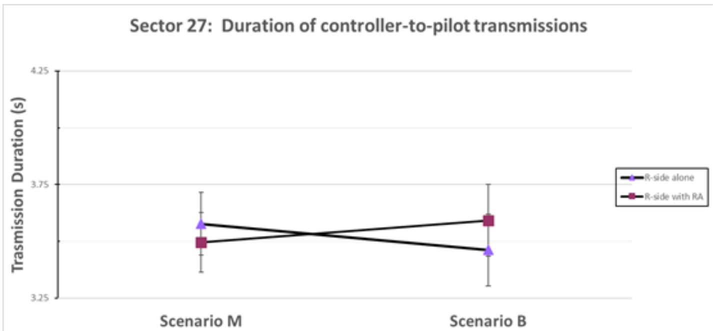


Figure 32. Sector 27: Average duration of controller-to-pilot transmissions

We also found a significant interaction of Scenario x Team Configuration on the duration of controller-to-pilot transmissions, $F(1,14) = 5.017$, $p = .042$, partial $\eta^2 = .264$. As shown in Figure 32, when participants worked with an RA, they made longer transmissions in the B scenario than the M scenario ($p < .05$), but when working alone, they made longer transmissions in the M scenario than the B scenario ($p < .05$).

For Sector 27 pilot-to-controller transmissions, we found a significant effect of Scenario on the number of transmissions, $F(1,14) = 103.38$, $p < .001$, partial $\eta^2 = .881$, with more transmissions made in the B scenario ($M = 159.72$, $SD = 12.26$) than the M scenario ($M = 124.47$, $SD = 21.92$), a result we expected given the greater number of aircraft in the B scenario. We did not find any significant effects of the test conditions on pilot-to-controller transmission durations, which had an overall average duration of 3.124 s ($SD = 0.318$).

5.1.5 Losses of Separation

We examined the instances in which a loss of separation (LOS) occurred using 5 nmi (9.26 km) horizontal and 1,000 ft (304.8 m) vertical separation standards. We eliminated LOS that were shorter than a single sweep of the radar (12 s) because the participants would not have been able to detect changes in aircraft position between radar updates. We also eliminated LOS that occurred because of a system error, simulation pilot error, or because of other factors not attributable to the participant. Upon review, we found that the majority of the LOS occurred between an aircraft in the participant’s sector and an aircraft in an adjacent sector. Because the only staffed sectors in our simulation were the ones controlled by the participants, it was not possible for them to coordinate with any other sectors to resolve potential conflicts. Therefore, we eliminated LOS that occurred with an aircraft in an unstaffed sector.

Only four LOS remained in the entire simulation, and therefore, it was not possible to conduct any statistical analyses. Instead, we provide a narrative description of these events. The four remaining LOS occurred for one of the participant teams during the first four test scenarios they completed. Three of the LOS occurred for the higher experience participants. All of the LOS occurred while the participants were working as R-sides alone. Three of them occurred during B scenarios, and two of those three occurred during the same scenario at about the same time, midway through the scenario.

In two of the four LOS, both pairs of aircraft involved were in Sector 27. In both of these instances, the participant cleared the aircraft in each pair to the same altitude and initially had a J-ring around one of the aircraft in the pair, suggesting that he was monitoring for separation. In one instance, the controller had the J-ring on the lead of two eastbound aircraft. He cleared one aircraft to FL180 by voice and the other to FL180 by Data Comm. This resulted in a loss of separation near the boundary. He attempted to resolve the conflict by speeding up the lead aircraft just as it was leaving the sector. In the second instance, both aircraft were heading north, both at FL210, parallel to one another with one aircraft a few nmi ahead. The controller had a J-ring around one of the aircraft, again suggesting he was monitoring for separation. The controller issued a “Direct to” clearance to the aircraft ahead, which caused it to turn into the path of the other aircraft. During the audio/video review of these events, the researchers heard the participant indicate that he was having difficulty keeping up with the scenario.

The other two LOS involved one aircraft in each of the staffed sectors but for which the participants did not coordinate. In one LOS, the participant in Sector 27 made a CTCC request to handoff the aircraft to Sector 10 at FL250. The Sector 10 participant did not respond to the request and that aircraft went into conflict with another already in Sector 10. During the audio/video review, the researchers heard the Sector 10 participant express surprise at realizing that he had responsibility for the aircraft that had been in his sector.

In the second LOS that occurred between Sector 10 and Sector 27, both aircraft were initially in Sector 27 prior to the LOS. One of the aircraft was eastbound at FL 240 and the other was northbound climbing from FL180 to FL210. The aircraft handed off to Sector 10 and remained at FL210. Shortly thereafter the participant in Sector 27 descended the eastbound aircraft from FL240 to FL180 and the two aircraft went into conflict with one another. The researchers overheard both of the participants expressing difficulty during this segment of the scenario. We also overheard the participants starting and stopping clearances (i.e., disregard), expressing uncertainty about aircraft locations for aircraft checking on, and being unsure about which aircraft was making a request.

5.1.6 Aircraft Deviations

We examined instances in which the participants did not manage the aircraft according to the airspace procedures. We evaluated when the participants did not initiate a handoff or did not issue a point out to the next sector if the aircraft was close to the boundary, did not transfer the voice frequency to the next sector, or did not transfer Data Comm eligibility to the next sector. These instances happened infrequently. We did not analyze these data via ANOVA because they consisted of a small number of frequency counts which are not normally distributed. Instead, we used other forms of multiple regression to analyze these count-type data. The logic of these analyses and the outputs are analogous to an ANOVA, but some of the statistical assumptions of the underlying models differ. The generalized linear model (GLM) is a class of regression models that allow researchers to match the underlying assumptions of the source data. We used a Poisson distribution model that reflects the underlying distribution of the count-type data from most of our deviation measures. We tested the statistical significance of the modeled effects of independent variables using an Analysis of Deviance (ANODE) approach which is analogous to an ANOVA but is based on a chi-squared distribution, not an F-distribution. The GLM was implemented in R software (Faraway, 2016) (Mangiafico, 2016). The models included the same three factors used in the ANOVAs for our other analyses: Experience Level, Scenario, and Team Configuration. We performed these analyses separately for Sector 10 and Sector 27, as we did for all other analyses, and only report statistically significant differences ($p < .05$).

5.1.6.1 Failure to transfer Data Comm eligibility

Overall, we found very few instances in which the participants failed to transfer Data Comm eligibility. For Sector 10, we found a mean of 2.59 failures to transfer Data Comm ($SD = 1.41$) overall. We did not find significant effects of Experience Level, Scenario, or Team Configuration.

For Sector 27, we found that the participants failed to transfer Data Comm more often when they worked as R-sides alone ($M = 5.31$, $SD = 2.24$) than when they worked with an RA ($M = 4.13$, $SD = 2.06$), $X^2(1) = 4.80$, $p = .029$. However, we found a significant interaction of Team Configuration and Scenario that indicated that the participants failed to transfer Data Comm eligibility less often when they worked the M scenario with an RA than they did when they worked the B scenario with or without an RA or when they worked the M scenario alone, $X^2(1) = 4.32$, $p = .038$ (see Table 3).

Table 3. Mean number of failures to transfer Data Comm eligibility in Sector 27

	M Scenario	**B Scenario**
R-side alone	5.38 (*SD* = 2.25)	5.25 (*SD* = 2.30)
R-side with RA	3.19 (*SD* = 1.56)	5.06 (*SD* = 2.11

5.1.6.2 Failure to transfer voice frequency

The number of instances in which the participants failed to transfer the frequency to the next sector was fairly low. We analyzed these data the same way we analyzed failures to transfer Data Comm eligibility. For Sector 10, there were more failures to transfer the frequency in the B Scenario (M = 12.79, SD = 3.92) than the M scenario (M = 9.91, SD = 3.17), X^2 (1) = 11.69, p < .001.

For Sector 27, there were also more failures to transfer the frequency in the B scenario (M = 14.94, SD = 5.57) than the M scenario (M = 10.28, SD = 4.61), X^2 (1) = 27.67, p < .001. We also found a main effect of Team Configuration, with more failures to transfer the frequency occurring when the participants worked as R-sides alone (M = 13.53, SD = 6.31) than when they worked with an RA (M = 11.69, SD = 4.67), X^2 (1) = 4.32, p = .038.

5.1.6.3 Failure to handoff aircraft or issue point out

We also evaluated the number of instances in which an aircraft got within 5 nmi of the sector boundary without the participant having started a handoff or issuing a point out. In Sector 10, there were more failures to handoff aircraft or issue point outs in the B scenario (M = 23.34, SD = 4.66) than in the M scenario (M = 20.69, SD = 2.87), X^2 (1) = 5.13, p = .023. In Sector 27, there was an average of 24.02 (SD = 4.68) failures to handoff or issue a point out, but these data did not vary significantly by test condition.

5.1.6.4 Airspace deviation summary

Overall, we did not find many airspace deviations. However, the results indicated that the participants committed more of them when they worked the B scenario and/or when working as an R-side alone. They made fewer deviations when they worked the M scenario or when working with an RA.

5.1.7 Command Modality

We evaluated whether the participants entered commands to the system differently across conditions by examining their use of keystroke and button press (point-and-click, trackball) entries. Although individual keystrokes far outnumbered individual button presses and button press entries took longer to execute than individual keystroke entries, we evaluated whether there was a shift in that pattern between the test conditions which could suggest that the participants were interacting with the system differently in different circumstances.

We evaluated the differences between the number of keystrokes made and the number of button presses made in each condition. In this analysis, the more positive the value, the more keystroke entries outnumbered button press entries. For Sector 10, the entries did not differ by test condition ($p > .05$). However, for Sector 27, we found that the difference between the number of keystroke entries and button press entries differed significantly by Team Configuration, $F(1,14) = 5.97$, $p = .028$, partial $\eta^2 = .30$. Keystrokes outnumbered button press entries more when participants worked alone ($M = 2832.21$, $SD = 1136.30$) than when they worked with an RA ($M = 2361.03$, $SD = 1057.11$).

We also examined the extent to which the durations of entries for button presses and keystrokes differed across conditions. In this analysis, the more positive the value, the longer button press entries took compared to keystroke entries. For Sector 10, we found that the differences in the durations of entries differed significantly by Scenario, $F(1,14) = 8.63$, $p = .011$, partial $\eta^2 = .38$. The durations of button press entries took significantly longer than keypress entries in the M scenario (M =1374.989, SD =1223.44) than in the B scenario (M = 871.08, SD=840.25). We did not find significant differences in the duration of entries for Sector 27 ($p > .05$).

Overall, the results suggest that the participants entered commands somewhat differently depending on the test condition, indicating that they may have opted to use a “faster” method of entry (keystrokes) when working as R-sides alone and “longer” method of entry (button presses) when working moderate—compared to busy—traffic scenarios.

5.2 Eye Tracking and Electroencephalography Data

The participants wore the EEG equipment and had their eye movements monitored when they worked the R-side positions at Sector 10 and Sector 27 in the test scenarios. We analyzed these data to obtain possible neurophysiological measures of workload and to investigate whether we would find common brain region activity associated with higher levels of display complexity and workload or patterns of activity that changed as a function of test condition.

5.2.1 Eye Tracking Data

We wanted to determine whether we would find indicators of cognitive workload via pupil diameter, as had been found in previous air traffic simulations (Ahlstrom & Friedman-Berg, 2006). We correlated pupil diameter with complexity of the FDB to determine whether we could find evidence that the amount of information presented was associated with a physiological manifestation of workload. We quantified FDB complexity by counting the number of fields occupied in the FDB. Complexity levels in the data ranged from 9 to 21, with the majority of the FDBs consisting of a complexity level between 12 and 18.

We did not find a correlation between FDB complexity and pupil diameter when we examined individual fixations. However, we also evaluated pupil diameter as a function of test condition to determine whether we would find any differences in workload as a function of Experience Level, Scenario, and Team Configuration. For Sector 10, we found a significant effect of Scenario, $F(1,14) = 5.14$, $p = .0397$, partial $\eta^2 = .269$ on pupil diameter (mm) with larger average pupil diameters found when the participants worked the B scenario ($M = 4.95$, $SD = 0.65$) than when they worked the M scenario ($M = 4.85$, $SD = 0.71$). Likewise, for Sector 27, we found a significant effect of Scenario on pupil diameter, $F(1,14)$ 8.97, $p = .0096$, partial $\eta^2 = .391$, with larger average diameters found when the participants worked the B scenario ($M = 5.45$, $SD = 0.64$) than when they worked the M scenario ($M = 4.83$, $SD = 0.711$). As we found for some of the air traffic data (and some of the subjective data reported later), the pupil diameter results indicated that the participants experienced higher workload when working the higher traffic scenarios.

5.2.2 EEG Data

We used the EEGLab toolbox for Matlab to reduce the raw EEG data into a format for data analysis (Delorme & Makeig, 2004). EEG data—like most physiological data—is complex and noisy and requires post-processing to increase the signal-to-noise ratio. We first applied a Butterworth band-pass filter between 0.1 and 30 Hz and removed DC offset drift from the EEG signal. Our next step was to visually inspect the data for channels and channel segments that contained bad data. Bad data can result from loosening of electrodes, excessive motion, and muscle artifacts. In our approach, we replaced data from bad channels with interpolated data from other surrounding electrodes whenever possible. We replaced a total of four bad channels from three of 128 recording sessions. Additionally, one of 128 recording sessions had bad data throughout and, therefore, we removed it from the analysis. Before the next analysis stage, the data were downsampled from 1000Hz to 240Hz to reduce computational load.

To analyze the EEG data, we created Fixation Event Related Potentials (fERPs). We recorded the onset of eye fixations from the SmartEye eye-tracker, limiting our analysis to fixations on FDBs. Those fixation times were imported into the EEG data and served as the events to which our fERPs were time-locked. We created data epochs of 250ms before and after each fixation. We filtered out fixations of less than 250ms in duration to prevent overlap of fixations during the epochs. We also excluded the first 3 minutes of each scenario while the screen was blank and aircraft were initially appearing on the participant displays.

After creating the data epochs, we further removed and reduced data artifacts. Our initial step was to remove trials containing voltage changes of ±1000mv. We then ran an Independent Component Analysis (ICA) to isolate components that account for the data variance due to artifacts from sources such as eye movements, eye blinks, and muscle movements. We ran the ADJUST algorithm (Mognon, Jovicich, Bruzzone, & Buiatti, 2011) to determine artifactual components, removed them from the data, and reconstructed the EEG signal without artifact contamination. We then removed any remaining artifactual epochs having microvolt variations of ±250mv.

We then used ICA to identify brain regions active during fixations on FDBs. ICA is a method for decomposing a multivariate signal into its likely independent components (ICs). Some of these ICs represent activity from an individual brain region. We localized the source of the ICs to their underlying brain regions. This process of source localization in EEG data is comparable to the process used to find sources of auditory noise or vibrations. By placing multiple sensors at known locations in or around the environment containing the sources (in the case of EEG, around the head and brain) and comparing the activity across the sensor locations, one can identify potential locations of the sources.

Our goal was to identify brain activity from regions common to all participants. We did this by concatenating the EEG data from all good sessions and running the ICA on the global set. We excluded the data from six of 128 sessions from analysis (one for bad data, and five for having less than 75 artifact-free trials). We then ran the ICA on the concatenated dataset, split the data back into individual sessions, and copied the global IC parameters to the individual datasets. We used the DIPFIT 2.x plugin for EEGlab to localize the global IC activity to likely electrical dipole generators within the brain. We set the parameters to give us dipoles that had less than 15% residual variance and constrained the search to regions within the brain.

5.2.2.1 Independent Component Analyses

We ran an ICA analysis on the EEG data to identify components that shared common variance of the participant's brainwaves. We used source localization to find the location in the brain from which that activity was likely originating. Our analysis revealed that during eye fixations of FDBs, four brain regions were active and produced electric signals that propagated to the scalp and our recording equipment. These four regions are not a complete list of all regions involved, since several other regions were likely active but not registered by our sensors or analysis. Our brief summary of the four regions we identified is provided in Table 4.

Table 4. Brain region and associated cognitive function

Location	Brain Region	Cognitive Functions
Talairach: (8, -57, 23) Brodmann Area 31	Right Posterior Cingulate Gyrus (PCG)	Part of the default network of Memory, Attention, and Decision-Making (Raichle, 2015) Learning and Change Detection (Pearson, Heilbronner, Barack, Hayden, & Platt, 2011)
Talairach: (-53, 1, -8) Brodmann Area 38	Left Anterior Temporal Pole (ATP)	Semantic Memory: knowledge of objects, people, words, and facts (Bonner & Price, 2013) Conceptual Object Processing- perception, evaluation, and use (Peelen & Caramazza, 2012)
Talairach: (60, -44, 29) Brodmann Area 39	Right Angular Gyrus (AG)	Semantic Processing: extracting meaning and reading comprehension (Seghier, 2013) Visuospatial Fact retrieval & calculations (Arsalidou & Taylor, 2011) Part of the default network of Memory, Attention, and Decision-Making (Mazoyer, et al., 2001)
Talairach: (36, -15, 38) Brodmann Area 4	Right Primary Motor Cortex (PMC)	Muscle Movements (Penfield & Boldrey, 1937) Motor Memory Consolidation (Meullbacher, et al., 2002)

We created Fixation Event Related Potentials (fERPs) to observe the neural activity of these brain regions when the participants fixated an FDB. To explore how these fERPs changed with respect to our test conditions, we compared the fERPs using an ANOVA. We averaged the data from all fixations in one experimental scenario to create a single fERP for each run. Each fERP epoch was 500 ms long (250 before and after fixation). For our analysis, we divided the epoch into ten 50 ms time windows, took the average amplitude of the fERP over that time window, and ran a separate 3-way ANOVA on the mean amplitude for each time window for the three independent variables in the study: Experience Level, Scenario, and Team Configuration. As we did for our other analyses, we analyzed the data for Sectors 10 and 27 separately. The sectors also differed from one another in that we recorded with two different EEG amplifiers and wireless transmitters. Analyzing them separately eliminated any potential differences in the data due to different equipment. Since the EEG data in this simulation involved a novel approach and data analysis, we had no a priori expectations about the fERP data. However, we hope that the data and the results will serve as a comparison for future studies.

There are some important considerations to be made before interpreting any fERP results. The first is that polarity—positive or negative—does not denote activity. A positive waveform does not mean the brain is becoming active nor does a negative waveform mean decreasing activity or inhibition. Imagine a simple electro-magnet with positive on one end and negative on the other. When you increase the activity of the magnet both positive and negative charge increase. If you buried that magnet inside the brain and turned it on, electrodes on one side of the head would read positive and those on opposite side would read negative. The positive and negative signals are resulting from the same "activity" inside the brain. The fERP component dipoles are very similar, with a positive and negative end. Changes in the waveform's amplitude represent changes in brain activity over time, but we cannot say this activity is increasing or decreasing. The absolute value of a single fERP is uninterpretable on its own. Analysis of fERP data focuses on differences in activity between conditions. An observed difference in the fERP waveforms from two conditions denotes a difference in brain activity in that condition. We do not know if the difference necessarily means more or less activity, simply that the neural activity in the brain region is different at a certain time under different conditions.

Table 5 presents the significant differences for the designated time windows resulting from our ANOVA ($p < .05$). All other comparisons revealed no significant differences.

Table 5. Results of fERP ANOVA

Brain Region	**Sector**	**Difference**	**Time Window**	***F*(1,53) and *p* values**
PCG	10	Experience	0-50ms post-fixation	$F = 7.46$, $p = .009$
PCG	27	Team Configuration	200-250ms post-fixation	$F = 4.60$, $p = .037$
AG	10	Team Configuration	50-100ms post-fixation	$F = 4.61$, $p = .036$
PMC	10	Scenario	200-150ms, 100-50ms, & 50-0ms pre-fixation	F = 5.14, 5.93, 4.17 p = .027, .018, .046

Figure 33 through Figure 36 present the fERPs that had a statistically significant difference between conditions in at least one of the ten 50-ms time windows. Each figure shows the fERPs from one sector and the relevant test conditions. The solid line is the average fERP and the colored shaded area surrounding it shows the 95% confidence interval for that data. The gray shaded bar denotes the time window for which the two fERPs were significantly different from one another.

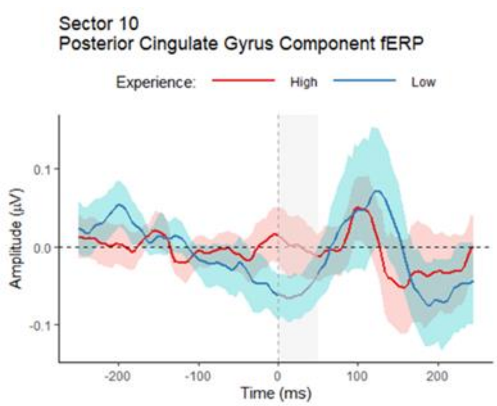


Figure 33. Sector 10: Posterior Cingulate Gyrus Component fERP

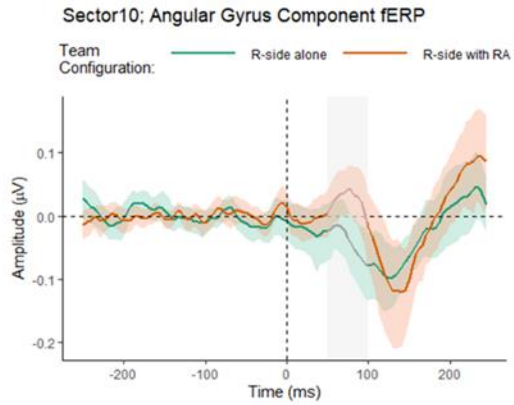


Figure 34. Sector 10: Angular Gyrus Component fERP

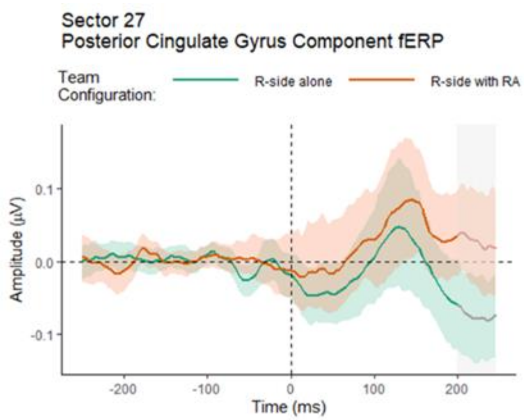


Figure 35. Sector 27: Posterior Cingulate Gyrus Component fERP

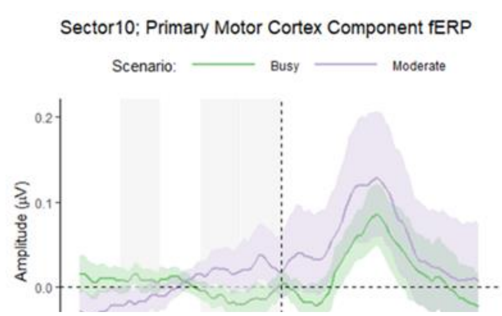


Figure 36. Sector 10: Primary Motor Cortex Component fERP

5.2.2.2 Data block complexity EEG analysis

We ran a second analysis on the EEG data we reduced from our initial ICA analysis. The previous analysis looked at the effects of the between-subject experiment IVs on the EEG data. This data block complexity analysis looked at the within-subject effect of FDB complexity on the EEG data. We used the same cleaned and reduced data from the ICA analysis. Each trial was coded as to the complexity of the FDB fixated upon. In this reduced dataset, FDBs had between 10 and 17 elements. There were not enough trials per participant to analyze each complexity level separately, so we combined 10, 11, and 12 into one bin and 15, 16, and 17 into another. This left us with four complexity levels for analysis: 12 or less, 13, 14, and 15+. We analyzed the fERPs for each of the four complexity levels for each of the four ICs corresponding to brain regions identified in the previous analysis. We also analyzed the two airspace sectors (10 and 27) separately, just as we did for all other analyses. We present the graphs of those fERPs in Figure 37 and Figure 38.

We analyzed the fERPs using a one-way repeated-measures ANOVA, with complexity level as the factor. We analyzed a variety of mean amplitude time windows along the ERP (e.g., 100–150ms). However, none of the comparisons between complexity fERPs in any of the comparisons at any time point were statistically significant at $p < .05$. Thus, our conclusion is that the brain activity from the four components we identified as active during fixation of FDBs did not differ according to FDB complexity.

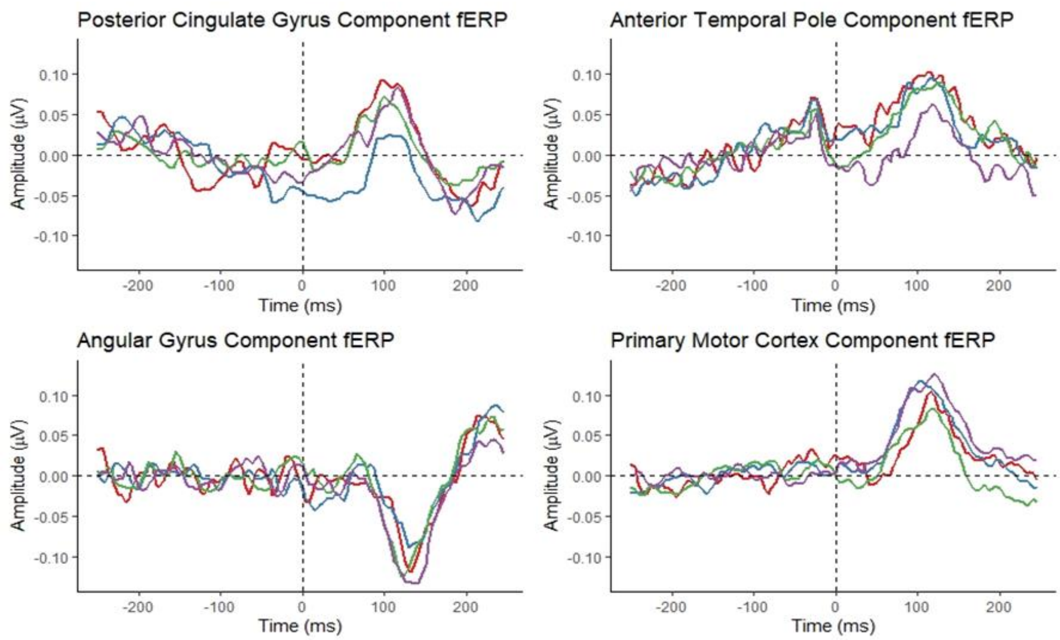


Figure 37. Sector 10: fERPs for FDB complexity level at each of four ICs.

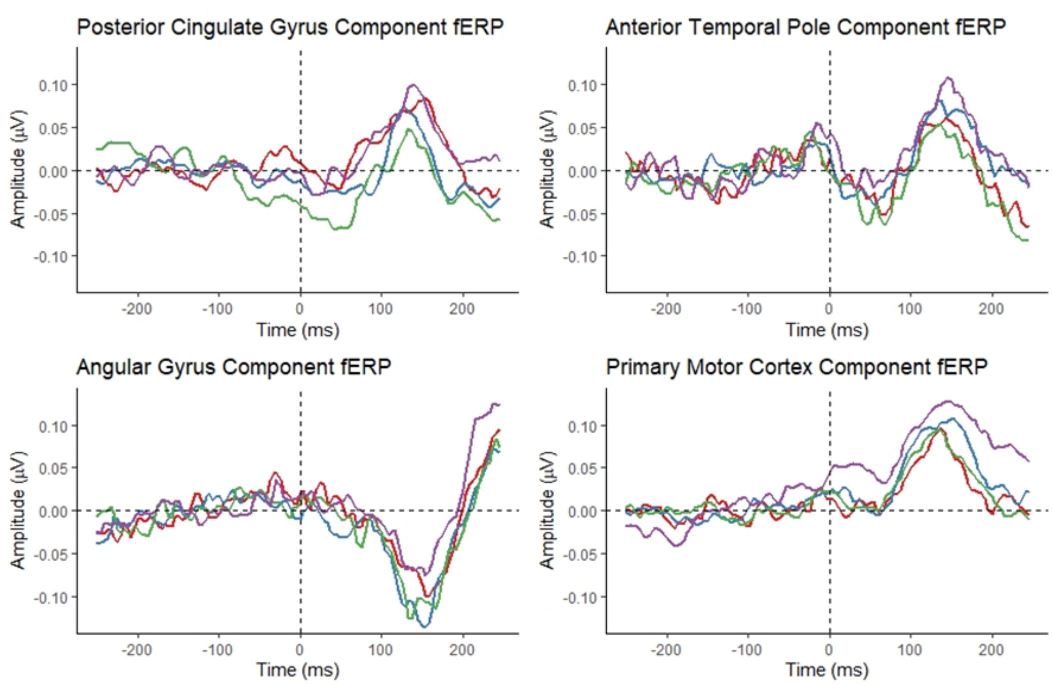


Figure 38. Sector 27: fERPs for FDB complexity level at each of four ICs

5.3 New Tool and Capability Use

We evaluated the number of interactions the participants made for each of the new tools and capabilities to obtain a measure of use. Overall, we found 85,977 total categorized interactions throughout all of the experimental runs. However, we did not categorize every interaction type due to time constraints. We focused on categorizing the interactions associated with each of the new tools: Conflict Probe, CTCC, TBFM, ABRR, and Data Comm. We considered an interaction to be any event in which the participant selected or made an entry for a particular tool via the trackball or keypad. For example, for Conflict Probe, we tallied an entry that displayed a route as one interaction and an entry that removed a displayed route as a separate interaction. We would not be able to readily determine if the participants were using the data provided by a tool to guide decisions that did not involve these direct interactions. For example, the participants could have used information provided by Conflict Probe data in the ACL to guide decisions on maneuvering aircraft without selecting the Conflict Probe icon to provide a graphic depiction of the potential conflict. Therefore, we implemented a definition of tool use that was directly observable.

Overall, new tool use comprised a low percentage of the total interactions (see Figure 39). Excluding Data Comm interactions, we found a total of 4808 interactions with the new tools, or 6%, of the total entries made across all experimental scenarios: 3310 interactions (69%) for CTCC, 721 interactions (15%) for ABRR, 651 interactions (14%) for TBFM, and126 interactions (3%) for Conflict Probe. For Data Comm, we found 3756 total interactions (4%), 899 (24%) of which were made when interacting with the other new tools.

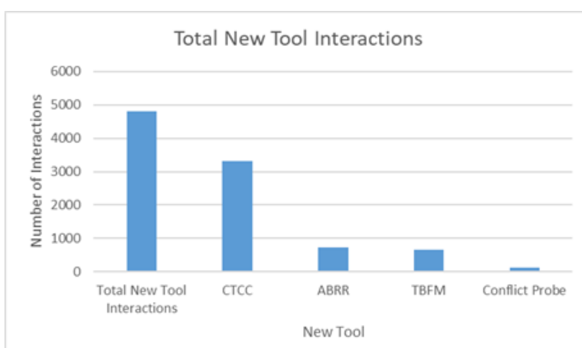


Figure 39. Number of new tool interactions

When working alone, the R-side participant interactions were made almost entirely from the R-side position. When working in R-side/RA teams, the number of new tool interactions was more evenly distributed between the R-side and RA participant for each sector as shown in Table 6.

Table 6. Number of new tool interactions by participant and team configuration

Position Making Entry					
Position	ZNY 10_ R-SIDE	ZNY_10_ RA	ZNY_27_ R-SIDE	ZNY_27_ RA	Grand Total
R-side alone	1219	9	827	2	2057
R-side w/RA	797	833	566	555	2751
Total	2016	842	1393	557	4808

We conducted an ANOVA to analyze the use of most of the new tools combined (Conflict Probe + CTCC + TBFM + ABRR). We tested the extent to which the new tools were used by the R-side participant as a function of Experience Level, Scenario, or Team Configuration. We tested sectors 10 and 27 separately. We analyzed the Data Comm interactions separately because Data Comm could be used when interacting with the new tools as well as with other types of commands.

As the data in Table 6 indicate, we found a significant effect of Team Configuration for both Sector 10 and 27. For Sector 10, the R-side participants interacted more with the new tools when working alone (M = 38.09, SD = 14.12) than when working with an RA (M = 22.78, SD = 14.12), $F(1,14)$ = 16.04, p = .001, partial η^2 = .539. Likewise, for Sector 27, the R-side participants interacted more with the new tools when working alone (M = 25.84, SD = 11.43) than when working with an RA (M = 17.34, SD = 8.01), $F(1,14)$ = 12.36, p = .003, partial η^2 = .469.

5.3.1 Data Comm

The participants used Data Comm to issue clearances for both the new tools as well as for other commands. We found a total of 3756 interactions for Data Comm for all commands, 899 of which were made when using the new tools. Almost all Data Comm interactions occurred via the R-side position in conditions when the R-side was working alone, but when the participants worked in teams, the number of Data Comm interactions from the RA position increased. The participants at the RA position were responsible for over 42% of the Data Comm interactions when the participants worked in teams. We provide the mean number of interactions for Data Comm by position in Table 7.

Table 7. Data Comm: Mean number (standard deviation) of interactions

	Sector 10				Sector 27			
R-side alone	**M**		**B**		**M**		**B**	
	R-side	**RA**	**R-side**	**RA**	**R-side**	**RA**	**R-side**	**RA**
L	38.875 (7.376)	0	28.75 (13.698)	0	21.125 (11.581)	0	8.375 (6.523)	0
H	31.375 (14.667)	0.25 (.707)	32.125 (16.711)	0	23.125 (17.033)	0.125 (.354)	14.125 (10.589)	0
R-side with RA								
L	26.25 (9.528)	20.875 (6.833)	26.125 (14.370)	26.375 (8.55)	18.50 (11.071)	12.25 (5.946)	8.25 (4.496)	10.5 (4.69)
H	25.625 (12.397)	15.25 (6.27)	21.5 (10.836)	18.125 (6.45)	17.75 (14.993)	4.75 (3.284)	10.25 (10.607)	5.25 (4.59)

We conducted an ANOVA to analyze the data. In Sector 10, the R-side participants interacted with Data Comm more when working alone (M = 35.67, SD = 10.36) than when working as part of a team, (M = 24.63, SD = 11.54), $F(1,13) = 25.66$, $p < .001$, partial $\eta^2 = .664$. For Sector 27, we found a significant effect of Scenario in which the R-side participants interacted with Data Comm more when working the M scenario (M = 20.13, SD = 13.20) than when working the B scenario (M = 10.25, SD = 7.67), $F(1,14) = 27.92$, $p < .001$, partial $\eta^2 = .666$.

5.3.2 Conflict Probe

The participants used Conflict Probe on the R-side infrequently. During the test scenarios, only eight of the 16 participants used the tool at least once at any point during the experiment scenarios. Out of 128 total scenarios, the R-side participants used Conflict Probe in only 26 of them. Table 8 summarizes the number of participants in each experimental condition who used Conflict Probe on the R-side.

Table 8. Conflict Probe: Number of participants using the tool on the R-side

	Sector 10		Sector 27	
R-side alone	**M**	**B**	**M**	**B**
Low Experience	2	1	1	1
High Experience	3	4	2	1
R-side with RA				
Low Experience	0	1	1	3
High Experience	3	2	1	0

We also found infrequent use of the tool from the RA position. Only eight (of 16) of the participants used Conflict Probe from the RA position at any point during their eight experimental scenarios. The participants at the RA position for both sectors interacted with Conflict Probe a total of 23 times through all of the experimental scenarios. Although FAA Order 7110.65 (Federal Aviation Administration, 2019) designates that the RA position should make use of such decision support tools when available, it is not known to what extent the RA uses Conflict Probe in the field. It would be useful to gather that data and compare it to what we observed in this simulation.

We tallied the number of interactions made from the R-side and RA positions in the experimental conditions and provide the summaries in Table 9. Given that few participants used Conflict Probe and the low number of interactions found for this tool, it was not possible to run any meaningful statistical analyses on these data, so we report only the raw number of interactions.

Table 9. Conflict Probe: Total number of interactions from R-side and RA positions

	Sector 10				Sector 27			
R-side alone	**Scenario M**		**Scenario B**		**Scenario M**		**Scenario B**	
	R-side	RA	R-side	RA	R-side	RA	R-side	RA
L	17	0	2	0	7	0	11	0
H	7	0	6	0	7	0	1	0
R-side with RA								
L	1	6	4	2	1	3	4	0
H	28	4	4	4	3	2	0	2
Totals	**53**	**10**	**16**	**6**	**18**	**5**	**16**	**2**

5.3.3 Time-Based Flow Management

TBFM was only available in Sector 27. Unlike Conflict Probe, all of the participants interacted with TBFM at least once during the experiment, even if they did not use it in every scenario. Most of the TBFM interactions (95%) were made by the R-side participants. Overall, 70% of the interactions, from either position, involved rejecting the speed or path stretch advisory recommended by the software. Table 10 summarizes the mean number of interactions with TBFM (and mean number of accepted and rejected advisories) in each of the experimental conditions for each position.

Table 10. TBFM: Mean number (standard deviation) of interactions from R-side and RA

	Sector 27							
R-side alone	**Scenario M**				**Scneario B**			
	R-side Position		RA Position		R-side Position		RA Position	
	Total	Accept	Total	Accept	Total	Accept	Total	Accept
L	9.5 (4.036)	2.625 (3.962)	0	0	8.5 (3.928)	2.625 (2.424)	0	0
H	11.75 (6.453)	5.125 (5.296)	0	0	9.5 (4.567)	1.625 (2.20)	0	0
R-side with RA								
L	8.375 (3.067)	1.25 (1.339)	1.5 (1.419)	0.375 (0.578)	8.625 (5.097)	2.75 (3.955)	1.625 (1.685)	0.625 (1.061)
H	9.75 (3.919)	4.875 (3.271)	1.0 (0.756)	0.25 (0.707)	8.875 (3.871)	2.375 (2.264)	0.375 (0.518)	0

Due to the low number of TBFM interactions, we analyzed the data using multiple Poisson regression implemented with the GLM function in R software (Faraway, 2016) (Mangiafico, 2016). We used the 3 experimental IVs as factors in the model and all interactions. We ran separate analyses for Sectors 10 and 27. Poisson regression is especially useful for data of “count” type with low numbers of trials. The significance of the factors was tested with an Analysis of Deviance (ANODE) approach, analogous to ANOVA. We did not find any significant effects of the test conditions, Experience Level, Scenario, Team Configuration, or any interactions. Thus TBFM use did not change due to our different experimental conditions.

5.3.4 Airborne Reroute

The ABRR interactions occurred almost exclusively on the R-side position when the participants worked as R-sides alone. Occasionally, the R-side participants used the RA-position interface to interact with ABRR and these were included in our analysis. The participants at the RA position made most of the ABRR interactions when the participants worked as R-side/RA team in Sector 10, $t(31) = 7.87$, $p < .001$, and Sector 27, $t(31) = 5.08$, $p < .001$. We summarize the mean number of ABRR interactions in the test conditions from each position in Table 11.

Table 11. ABRR: Mean number (standard deviation) of interactions from R-side and RA

	Sector 10				**Sector 27**			
R-side alone	**M**		**B**		**M**		**B**	
	R-side	**RA**	**R-side**	**RA**	**R-side**	**RA**	**R-side**	**RA**
L	6.5 (3.545)	0	4.625 (4.438)	0	3.5 (1.927)	0	0.875 (0.991)	0
H	3.625 (3.159)	1.125 (2.232)	3.875 (4.734)	0	4.875 (3.681)	0.25 (0.707)	1.875 (3.182)	0
R-side with RA								
L	2.5 (2.878)	7.0 (4.071)	2.125 (2.696)	7.25 (4.33)	1.25 (1.488)	4.875 (0.99)	1.375 (3.114)	5.0 (2.976)
H	1.5 (1.773)	5.25 (2.252)	2.125 (1.885)	7.5 (2.507)	2.75 (2.659)	3.5 (2.07)	1.625 (2.387)	3.0 (2.07)

We analyzed ABRR use using multiple Poisson regression in the same way as TBFM. For Sector 10, participants in the Low Experience group used ABRR more often ($M = 3.94$, $SD = 3.704$) than the participants in the High Experience group ($M = 2.78$, $SD = 3.14$), $X^2(1) = 6.399$, $p = .011$. We also found a significant effect of Team Configuration in which the participants interacted with ABRR more when working as R-sides alone ($M = 4.66$, $SD = 3.99$) than when working with an RA ($M = 2.06$, $SD = 2.27$), $X^2(1) = 32.889$, $p < .001$. Although the R-side controllers used ABRR more when working alone, ABRR was used more often by the R/RA team together than the R-side alone, $t(31) = 6.20$, $p < .001$.

For Sector 27, we found many effects of our IVs. The participants in the High Experience group used ABRR more often ($M = 2.78$, $SD = 2.97$) than the participants in the Low Experience group ($M = 1.75$, $SD = 2.20$), $X^2(1) = 7.58$, $p = .006$, the opposite of what we found for Sector 10. We found an interaction between Scenario and Team Configuration, $X^2(1) = 5.17$, $p = .023$, means shown in Table 12. We tested the simple effects by adding posthoc comparisons to our regression model, with a Tukey correction for multiple comparisons. We found a significant effect of Team Configuration in which the participants interacted with ABRR more when working as R-sides alone, but only in the M scenario, $p < .001$. The R-side controllers also used ABRR more in the M scenario than the B scenario, but only when working alone, $p < .001$. Although the R-side participants used ABRR more when working alone in the M scenario, the R-side/RA teams used ABRR more often than the R-sides alone for both scenarios, $t(31) = 3.90$, $p < .001$.

Table 12. ABRR: Mean number (standard deviation) of interactions for Sector 27

	M Scenario	**B Scenario**
R-side alone	4.19 (*SD* = 2.93)	1.38 (*SD* = 2.34)
R-side with RA	2.0 (*SD* = 2.22)	1.5 (*SD* = 2.68)

5.3.5 Controller-to-Controller Coordination

We found a greater number of CTCC interactions than TBFM, ABRR, and Conflict Probe interactions. There were many different types of opportunities to use CTCC, such as to propose conditions at handoff for an aircraft entering the sector (e.g., altitude), approving or denying requests, modifying proposed requests, or canceling requests. Each of these could involve several interactions with the tool. We present the mean number of interactions for CTCC by the R-side and RA positions for each of the test conditions in Table 13.

Table 13. CTCC: Mean number (standard deviation) of interactions from R-side and RA

	Sector 10				**Sector 27**			
R-side alone	**M**		**B**		**M**		**B**	
	R-side	**RA**	**R-side**	**RA**	**R-side**	**RA**	**R-side**	**RA**
L	33.375 (12.850)	0	28.750 (13.698)	0	11.875 (7.415)	0	12.875 (9.269)	0
H	35.75 (17.425)	0	32.125 (16.711)	0	11.625 (7.927)	0	13.375 (7.539)	0
R-side with RA								
L	19.625 (18.181)	20.125 (13.325)	22.875 (19.780)	10.375 (12.345)	5.25 (4.832)	12.125 (9.48)	6.125 (5.303)	13.625 (10.211)
H	18.125 (11.922)	15.75 (13.350)	17.875 (12.017)	18.875 (13.314)	7.125 (5.842)	6.375 (8.366)	7.375 (6.989)	13.125 (12.124)

We conducted an ANOVA to analyze the data. We found a significant effect of Team Configuration for both Sector 10 and Sector 27. For Sector 10, the R-side participants interacted more with the new tools when working alone ($M = 32.50$, $SD = 14.70$) than when working with an RA ($M = 19.63$, $SD = 15.45$), $F(1,14) = 10.44$, $p = .006$, partial $\eta^2 = .427$. Although the R-side controllers used CTCC more when working alone, CTCC was used more often by the R/RA team together than the R-side alone, $t(31) = 2.78$, $p = .005$. This result suggests that the environment was likely too busy for the R-side participants to make use of CTCC on their own and that they relied more on the RA to use the tool when they worked as a team.

Likewise, for Sector 27, the R-side participants interacted more with CTCC when working alone ($M = 12.44$, $SD = 7.84$) than when working with an RA ($M = 6.47$, $SD = 5.61$), $F(1,14) = 16.63$, $p = .001$, partial $\eta^2 = .543$. Although the R-side controllers used CTCC more when working alone, CTCC was used more often by the R/RA team together than the R-side alone, $t(31) = 2.38$, $p = .012$.

We reviewed some of the new tool usage by viewing the simulation recordings to ensure that the entries were categorized correctly. During our review, it was apparent that some of the CTCC interactions were likely made unintentionally. We attributed this to the inconsistencies in the use of the PICK and ENTER trackball buttons. The PICK and ENTER trackball buttons open different menu options from the same area of the FDB. Figure 40 shows the three buttons available on the trackball. The participants had to remember to slew to the correct portal on the data block, such as speed or CID, and then to use the correct trackball button to bring up the specific option associated with that portal.

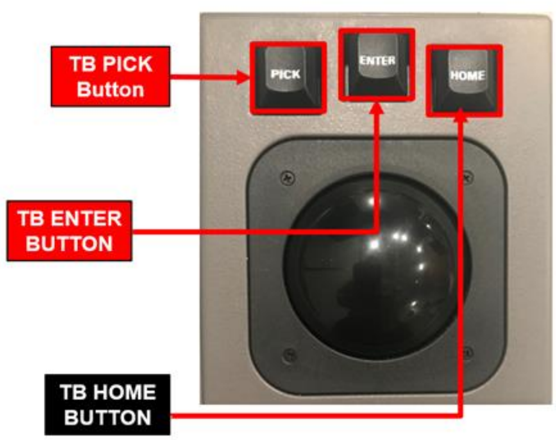


Figure 40. Trackball buttons

Slewing to the speed portal on the data block and then selecting the PICK button opens the speed fly-out menu. But, selecting the ENTER button on the speed portal opens the coordination menu. Selecting the HOME button does not do anything when accessed on the speed portal. Although we briefed the participants on the different button functions in training, it was likely that given their short experience with the tools, there were situations in which they mistakenly selected an option other than the one intended. In our review, we found several instances in which the participants brought up the speed menu by selecting the PICK key when they were intending to bring up the coordination menu by selecting the ENTER key or vice versa. Therefore, the number of interactions we counted in this simulation for CTCC likely included a number of instances in which the participant chose it unintentionally. This critical human factors issue was identified during Phase 1 of the OAIP project (Willems & Dworsky, 2018) and one that must be addressed going forward to ensure that controllers are accurately entering a desired command in a timely manner.

5.4 Subjective Data

The participants provided feedback about their performance and workload throughout the simulation. During each scenario, they provided workload ratings when prompted every two minutes via the WAK device. At the end of each test scenario, the participants completed the PSQ to provide feedback on various aspects of the scenario just completed. The WAK and PSQ data are both values on a 10-point Likert scale and are not normally distributed, violating the normality assumptions of an ANOVA model. Ordinal regression is a non-parametric approach that takes advantage of the fact that the dependent variable is ordinal and cumulatively distributed. Since each scenario had multiple WAK ratings and our dependent variable was an average of those ratings, a parametric model was appropriate since we had normally distributed residuals. However, the PSQ data was a single response per scenario and needed a non-parametric analysis. We used Multiple Ordinal Regression to analyze the PSQ data. The logic of this analysis and outputs are analogous to an Analysis of Variance (ANOVA) but the statistical assumptions of the underlying models differ. Ordinal regression is a non-parametric approach that takes advantage of the fact that the dependent variable is ordinal and cumulatively distributed. We tested for the significance of the effects of the independent variables using an analysis of deviance (ANODE) approach, which is analogous to an ANOVA. The ordinal regression was implemented in R software with a Cumulative Link Model (Christensen, 2015) (Mangiafico, 2016). The model included the three test conditions: participant Experience Level, Scenario, and Team Configuration. We performed the analyses for the WAK and PSQ data separately for the two airspace sectors (10 and 27) as we did for all of the data in this simulation.

5.4.1 Workload Assessment Keypad Responses

The participants entered their workload ratings (1 = Extremely Low; 10 = Extremely High) on the Workload Assessment Keypad (WAK) at 2-minute intervals throughout each scenario. Overall, the participants responded to 92% of the prompts. However, one group of participants responded to the vast majority (97%) of prompts with a rating of "1" so we eliminated that group's data from the analysis. All of the other groups' responses varied throughout the scenarios and included the full range of the scale.

We computed the average of the ratings to obtain an overall workload measure for each scenario. We analyzed the data for each sector separately using a 3-way ANOVA to determine whether there were significant differences between the test conditions.

For Sector 10, the average rating was 3.33 (SD = 1.36). The participants rated their workload higher when they worked as R-sides alone (M = 3.83, SD = 1.49) than when they worked working with an RA (M = 2.83, SD = 1.00), $F(1,40) = 9.46$, $p = .004$. The participants also rated their workload higher when they worked the B scenario (M = 4.01, SD = 1.35) than when they worked the M scenario (M = 2.65, SD = 0.98), $F(1,40) = 17.50$, $p < .001$.

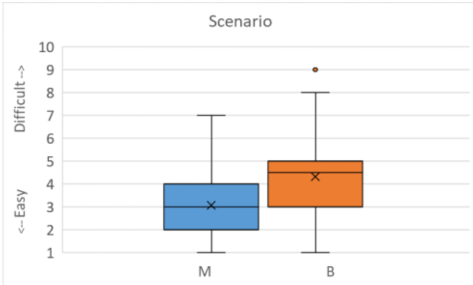
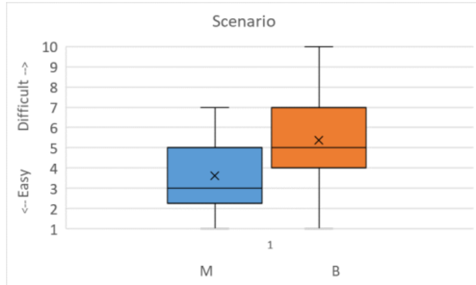
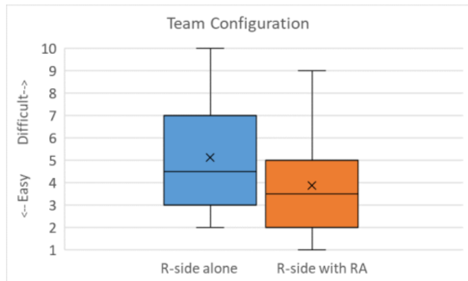
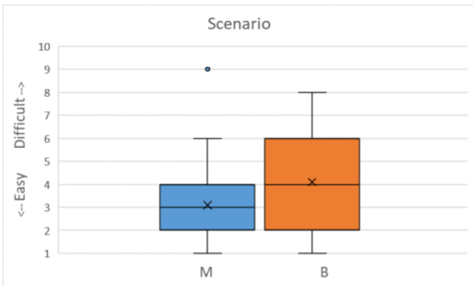
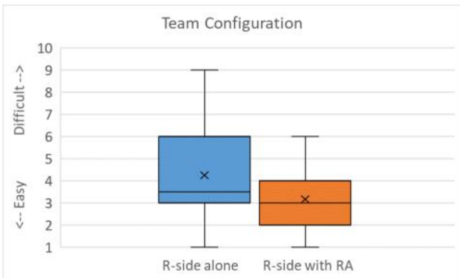
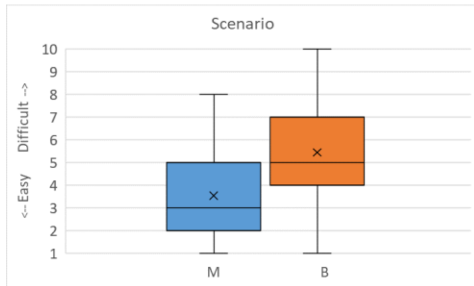
Sector 27 had a similar overall average rating of 3.55 (SD = 1.91) and followed the same pattern of results. We found the participants rated their workload higher when they worked as R-sides alone (M = 4.06, SD = 2.18) than when they worked with an RA (M = 3.03, SD = 1.48), $F(1,40) = 4.52$, $p = .040$. We also found that the participants rated their workload higher when they worked the B scenario (M = 4.52, SD = 1.81) than when they worked the M scenario (M = 2.57, SD = 1.49), $F(1,40) = 16.17$, $p < .001$.

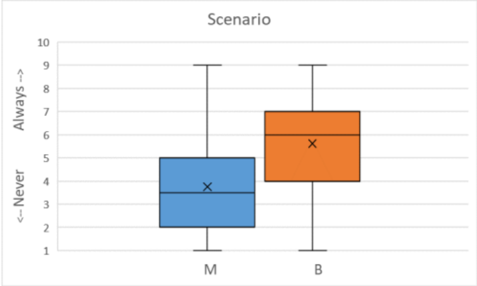
The participants also provided ratings of different types of workload for each scenario on the Post-Scenario Questionnaire (items 11, 12, and 13).

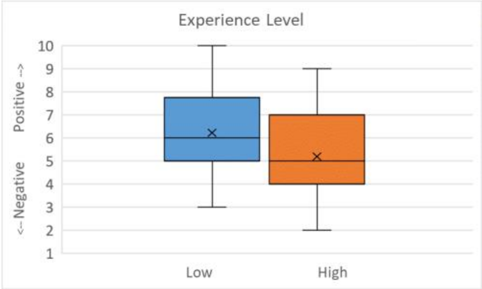
5.4.2 Post-Scenario Questionnaire Responses

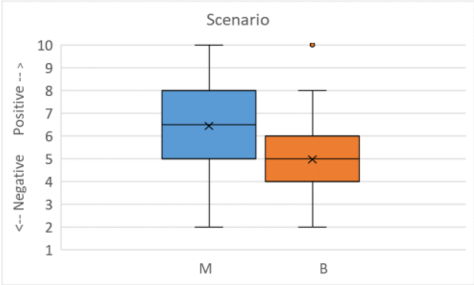
After the participants completed each test scenario, they completed the PSQ (seeappendix C). We analyzed the data from the PSQ using Multiple Ordinal Regression. We present the median responses, response ranges, and quartiles and indicate significant differences found for each sector in Table 14.

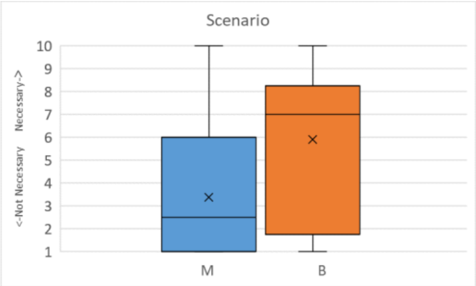
Table 14. PSQ responses

PSQ Item	Sector 10 IV: X2 (1), p = * <.05; ** <.01; *** <.001	Sector 27 IV; X2 (1); p = * <.05; ** <.01; *** <.001
1. Overall, how easy was it to use available information during this scenario run?	 Scenario B more difficult (X^2 = 8.22 **)	 Scenario B more difficult (X^2 = 12.64 ***)  Working alone more difficult (X^2 = 7.37 **)
2. How easy was it to view presented information during this scenario run?	 Scenario B more difficult (X^2 = 8.79 **)  Working alone more difficult (X^2 = 5.22 *)	 Scenario B more difficult (X^2 = 12.70 ***)

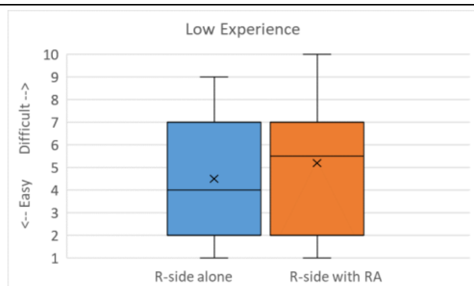
<table>
<tr><th>PSQ Item</th><th>Sector 10
IV: X2 (1), p =
* <.05; ** <.01; *** <.001</th><th>Sector 27
IV; X2 (1); p =
* <.05; ** <.01; *** <.001</th></tr>
<tr><td>3. How often did the presented information cause clutter on the display during this scenario run?</td><td>ns</td><td>

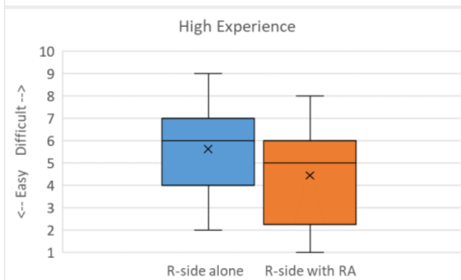
Scenario B more difficult (X^2 = 11.68 ***)</td></tr>
<tr><td>4. What effect did the presented information have on your ability to control traffic safely during this scenario run?</td><td>ns</td><td>

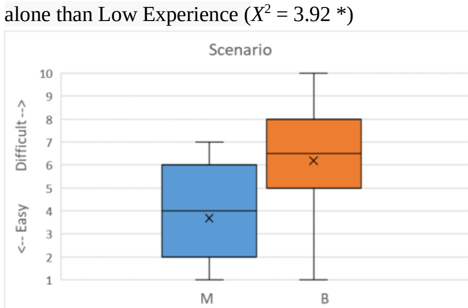
Low Experience rated a more positive effect (X^2 = 4.42 *)


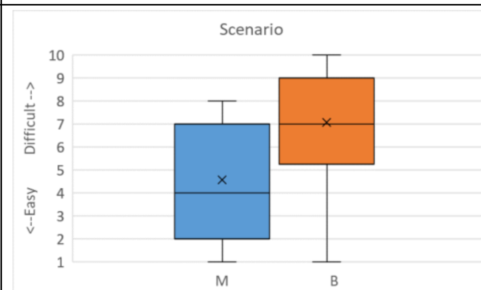
Scenario M had more positive effect (X^2 = 9.09 **)</td></tr>
<tr><td>5. Was the Radar Associate (RA) side being staffed essential to handle the traffic in this scenario?</td><td>

RA position rated more necessary when working Scenario B (X^2 = 9.28 **)</td><td>ns</td></tr>
<tr><td>6. Rate your level of Air/Ground communication</td><td>ns</td><td>ns</td></tr>
</table>

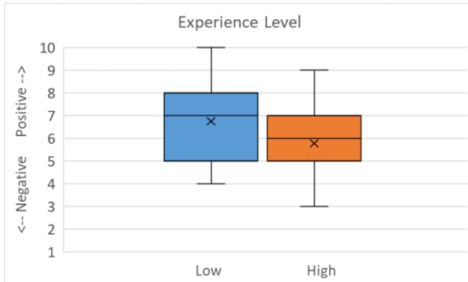
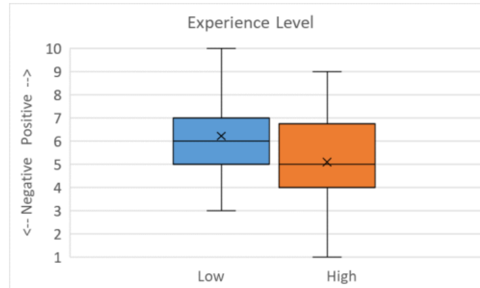
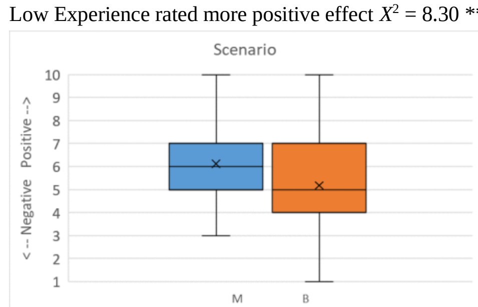
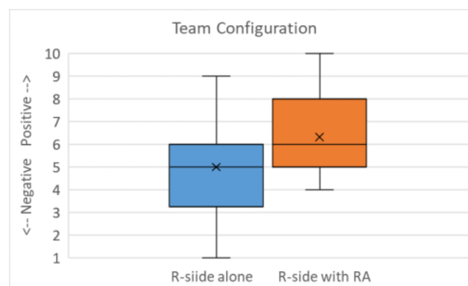
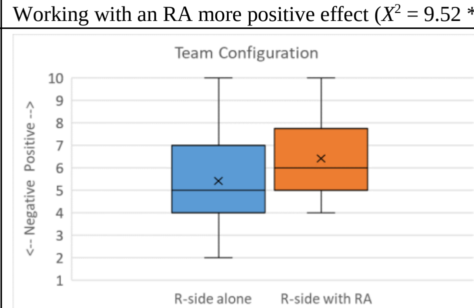
<table>
<tr><th>PSQ Item</th><th>Sector 10
IV: X2 (1), p =
* <.05; ** <.01; *** <.001</th><th>Sector 27
IV; X2 (1); p =
* <.05; ** <.01; *** <.001</th></tr>
<tr><td>7. Rate your level of situational awareness during this scenario.</td><td>ns</td><td>ns</td></tr>
<tr><td>8. Rate the performance of the simulation pilots in terms of their responding to control instructions and providing readbacks.</td><td>ns</td><td>ns</td></tr>
<tr><td>9. Rate the difficulty of this scenario.</td><td>

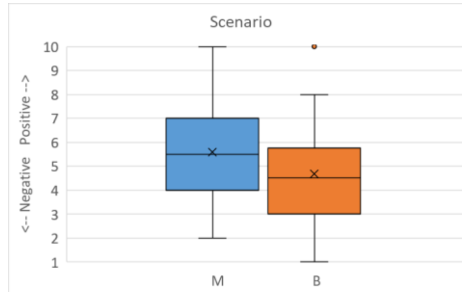
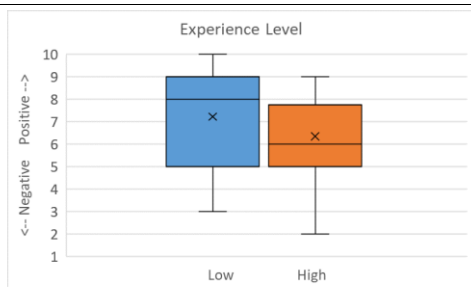
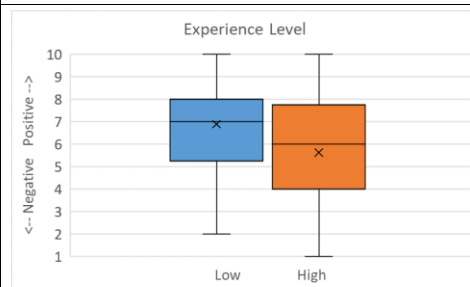
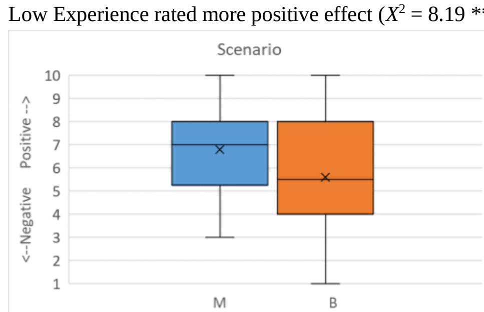
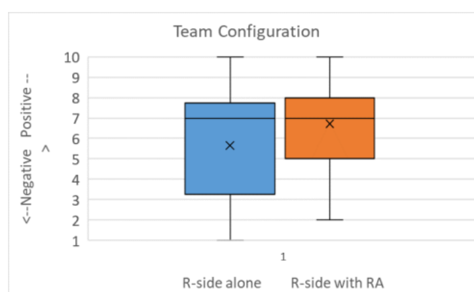


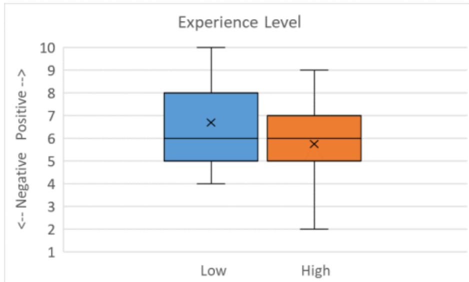
High Experience reported more difficulty working alone than Low Experience (X^2 = 3.92 *)


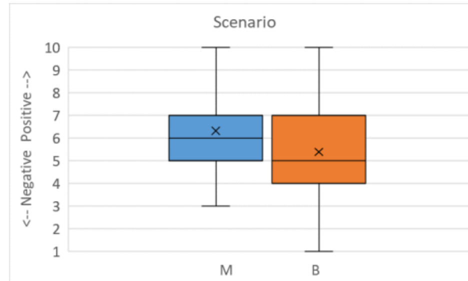
Scenario B more difficult (X^2 = 21.24 ***)</td><td>

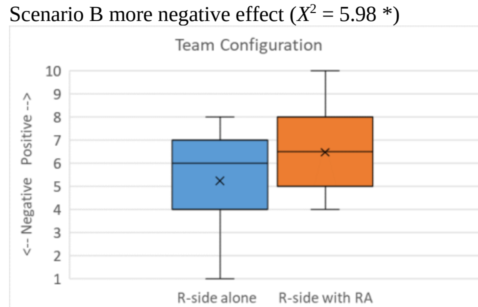
Scenario B more difficult (X^2 = 24.8156 ***)</td></tr>
<tr><td colspan="3">How did the amount of presented information affect your ability to perform the following tasks?</td></tr>
</table>

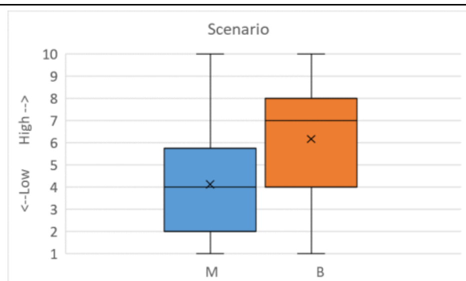
<table>
<tr><th>PSQ Item</th><th>Sector 10
IV: X2 (1), p =
* <.05; ** <.01; *** <.001</th><th>Sector 27
IV; X2 (1); p =
* <.05; ** <.01; *** <.001</th></tr>
<tr><td>10a. Situation monitoring</td><td>
Low Experience rated more positive effect X^2 = 5.29 *)</td><td>
Low Experience rated more positive effect X^2 = 8.30 **)

Scenario M more positive effect (X^2 = 4.39 *)

Working with an RA more positive effect (X^2 = 9.52 **)</td></tr>
<tr><td>10b. Searching for potential aircraft conflicts</td><td>ns</td><td>
Working with an RA more positive effect (X^2 = 4.34 *)</td></tr>
<tr><td>10c. Resolving potential conflicts</td><td>ns</td><td>ns</td></tr>
</table>

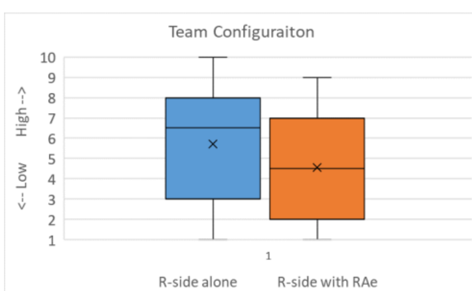
PSQ Item	Sector 10 IV: X2 (1), p = * <.05; ** <.01; *** <.001	Sector 27 IV; X2 (1); p = * <.05; ** <.01; *** <.001
10d. Managing air traffic sequences	ns	 Scenario B more negative effect (X^2 = 4.54 *)
10e. Re-routing or evaluating flight plans	 Low Experience rated more positive effect (X^2 = 4.60*)	 Low Experience rated more positive effect (X^2 = 8.19 **)  Scenario B more negative effect (X^2 = 4.33 *)  Working with an RA more positive effect (X^2 = 4.94 *)

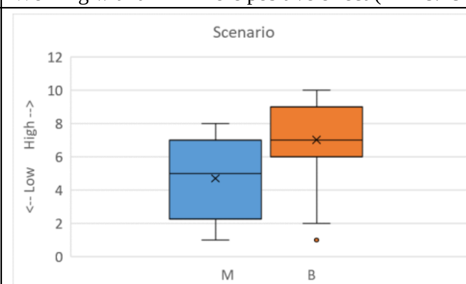
<table>
<tr><th>PSQ Item</th><th>Sector 10
IV: X2 (1), p =
* <.05; ** <.01; *** <.001</th><th>Sector 27
IV; X2 (1); p =
* <.05; ** <.01; *** <.001</th></tr>
<tr><td>10f. Managing sector/position resources</td><td>

Low Experience rated more positive effect (X^2 = 4.87*)</td><td>

Scenario B more negative effect (X^2 = 5.98 *)


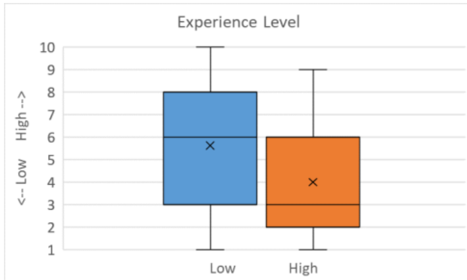
Working with an RA more positive effect (X^2 = 8.18 *)</td></tr>
<tr><td>11. Rate your mental demand during this scenario (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.).</td><td>

Scenario B more mentally demanding (X^2 = 10.69 **)


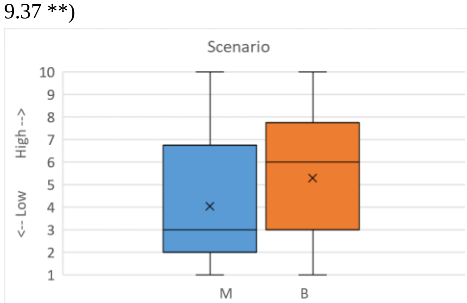
Working alone more mentally demanding (X^2 = 3.90 *)</td><td>

Scenario B more mentally demanding (X^2 = 13.42 **)</td></tr>
</table>

<table>
<tr><th>PSQ Item</th><th>Sector 10
IV: X2 (1), p =
* <.05; ** <.01; *** <.001</th><th>Sector 27
IV; X2 (1); p =
* <.05; ** <.01; *** <.001</th></tr>
<tr><td>12. Rate your physical demand during this scenario (e.g., communications and key presses).</td><td>

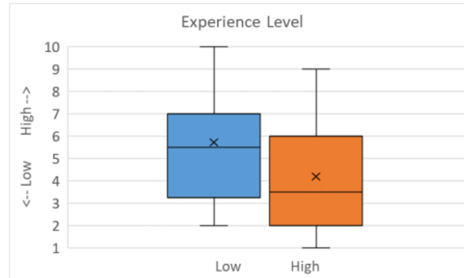


Low Experience rated more physical demand (X^2 = 9.37 **)

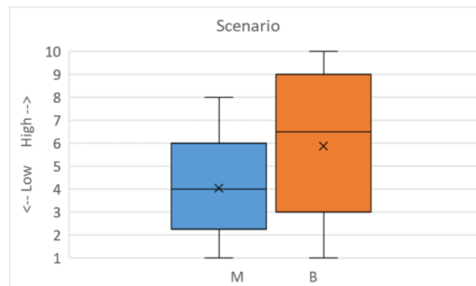


Scenario B more physically demanding (X^2 = 4.91 *)

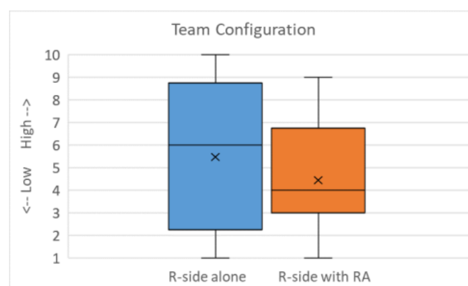
</td><td>



Low Experience rated more physical demand (X^2 = 7.99 **)

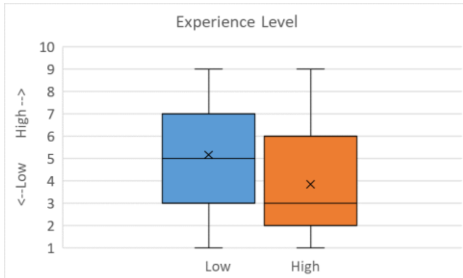
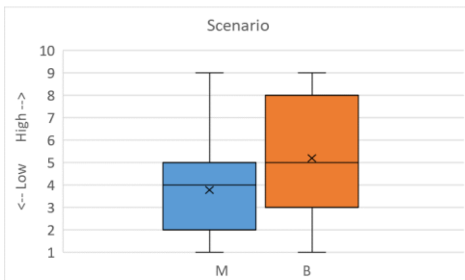
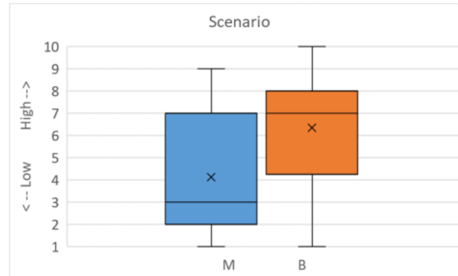


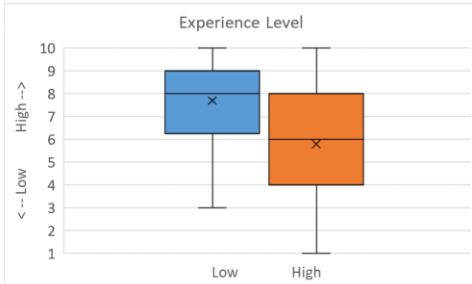
Scenario B more physically demanding (X^2 = 9.28 **)

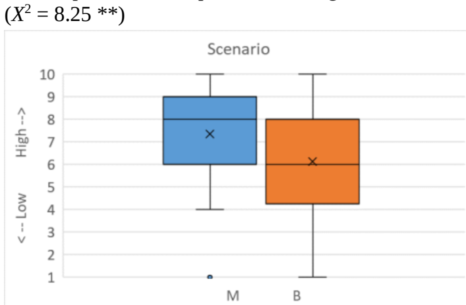


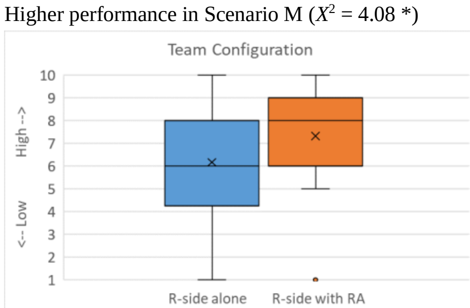
Working alone more physically demanding (X^2 = 4.23 *)

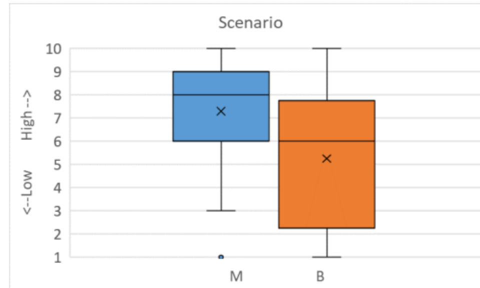
</td></tr>
</table>

PSQ Item	**Sector 10** **IV: X2 (1), p =** *** <.05; ** <.01; *** <.001**	**Sector 27** **IV; X2 (1); p =** *** <.05; ** <.01; *** <.001**
13. Rate your temporal demand during this scenario. (How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred?)	 Low Experience rated more temporal demand (X^2 = 4.79 *)  Scenario B more temporally demanding (X^2 = 6.42 *)	 Scenario B more temporally demanding (X^2 = 11.80 ***)

<table>
<tr><th>PSQ Item</th><th>Sector 10
IV: X2 (1), p =
* <.05; ** <.01; *** <.001</th><th>Sector 27
IV; X2 (1); p =
* <.05; ** <.01; *** <.001</th></tr>
<tr><td>14. Rate your performance during this scenario. (How successful do you think you were in accomplishing the goals of the task?)</td><td>

Low Experience rated performance higher (X^2 = 8.25 **)


Higher performance in Scenario M (X^2 = 4.08 *)


Higher performance with RA (X^2 = 4.87 *)</td><td>

Higher performance in Scenario M (X^2 = 10.82 **)</td></tr>
</table>

Overall, for both Sector 10 and Sector 27, the participants reported finding the B scenarios more difficult to manage than the M scenarios and that they had more difficulty managing the scenarios when working as R-sides alone than when working with an RA.

We found some differences with respect to Experience Level. The participants in the Low Experience group generally reported that the information presented during the scenarios had a more positive effect on their performance, control of traffic, situation monitoring, managing sector resources, and rerouting and evaluating flight plans than did the participants in the High experience group. However, the participants in the Low Experience group also reported greater effort and frustration, and greater physical and temporal demand in the scenarios than did the participants in the High Experience group.

5.4.3 Exit Questionnaire Responses

The participants completed the Exit Questionnaire at the end of the simulation (see appendix D). This questionnaire asked the participants about their experience in the simulation and included questions about simulation realism, the effectiveness of the training provided, and their reactions to the new capabilities, including the colors and symbols used, the effect on their ability to manage traffic, maintain situation awareness, and so forth, and the effect of the new symbols on display clutter.

The questions included 10-point scales for participants to provide ratings about each item from low (1) to high (10). We compared the responses of the participants in the Low Experience group with the participants in the High Experience group by conducting a Mann-Whitney non-parametric test because of the ordinal nature of the rating scale data and the assumption that these data were not normally distributed. We ran these tests on each of the questions on the Exit Questionnaire. The results are presented in Table 15, with significant differences between the groups indicated by the asterisk next to the bold, italicized question. Not all of the participants provided responses to all of the questions.

Table 15. PEQ responses. Significant differences are shown bold and italicized

Exit Questionnaire Category/Question	**Low Experience Median (range)**	**High Experience Median (range)**
Simulation Realism and Research Apparatus		
1. Rate the overall realism of the simulation.	5 (2 - 9)	5 (3 - 8)
2. Rate the realism of the simulation hardware compared to actual equipment.	6 (4 - 10)	8 (3 - 8)
3. Rate the realism of the simulation software compared to actual equipment.	6 (3 - 10)	6.5 (3 - 8)
4. Rate the realism of the simulation traffic scenarios compared to actual NAS traffic.	6.5 (3 - 8)	5 (2 - 8)
5. To what extent did the WAK workload rating technique interfere with your ATC performance?	3.5 (1 - 10)	3.5 (1 - 9)
6. Did you feel that there were enough practice scenarios to familiarize you with the new elements in ERAM?	9 (8 - 10)	9 (7 - 10)
Training		
1. DESIREE	7.5 (6 - 10)	8 (5 - 9)
2. Data Comm	8.5 (4 - 10)	8.5 (3 - 9)
3. Traffic Based Flow Management (TBFM)	7 (2 - 9)	6 (1 - 8)

Exit Questionnaire Category/Question	Low Experience Median (range)	High Experience Median (range)
4. Separation Management (SEPMAN)	6 (5 - 10)	5.5 (3 - 8)
Symbol Complexity		
1. Were you able to determine the transfer of communication status of each aircraft?	9 (4 - 10)	8.5 (3 - 10)
2. Could you determine the difference between equipped and non- equipped Data Comm aircraft.	9.5 (6 - 10)	10 (4 - 10)
3. Were you able to determine track ownership of the datablocks?	10 (7 - 10)	9 (3 - 10)
4. Could you tell which aircraft were ADS-B equipped?	3 (1 - 10)	4 (1 -9)
5. Were the safety alerts easy to see in the data blocks?	7 (5 - 10)	8 (3 - 10)
6. Could you determine the meaning of different colored symbols?	9 (5 - 9)	6.5 (1 - 9)
Traffic Management		
1. Was communications delay an issue during the simulation?	8 (4 - 10)	7 (4 - 10)
2. How difficult did you find the airborne reroute process?	2 (1 -7)	3.5 (1-8)
3. How helpful did you find the Delay Countdown Timer to controlling traffic?	4.5 (1 - 8)	5 (1 - 9)
Data Comm		
1. Did you find the Data Comm symbols confusing?	2 (1 - 6)	3.5 (1 - 10)
2. Were you able to determine when an aircraft was data comm equipped?	9.5 (8 - 10)	8.5 (4 - 10)
3. Did color help with your interpretation of Data Comm?	7.5 (3 - 10)	7 (3 - 10)
4. Did the increased datablock symbols cause clutter on your scope?	4 (1 - 7)	6 (2 - 10)
5. Could you determine the status of a message you sent to an aircraft?	8 (4 - 10)	7.5 (3 - 9)
6. How quickly were you able to send messages?	7.5 (3 - 10)	7.5 (3 - 10)
7. Did you feel forced to use one input method over another (i.e keyboard vs trackball)	6 (1 - 10)	5 (2 - 10)
8. Did Data Comm flow with the rest of the system?	7.5 (6 - 10)	7 (3 - 9)
9. Did you feel like you had to remember more in your head while using Data Comm?	3.5 (1 -7)	4.5 (2 - 9)
Time Based Flow Management (TBFM)		
1. *Did you find the TBFM symbols confusing?*	***1.5 (1 - 5)***	***6 (3 – 9)***
2. Were you able see indication that your traffic goal was achieved?	7 (2 - 10)	3 (1 - 7)
3. Did you find the extra data blocks distracting?	3 (1 - 4)	5 (2 - 10)
4. Did the colors used in the Range Data block confuse you?	3 (1 - 5)	4.5 (1 - 6)

Exit Questionnaire Category/Question	Low Experience Median (range)	High Experience Median (range)
5. How much did the Range Data Block and Delay Time Counter help you absorb delay time?	3.5 (1 - 8)	6 (1 - 8)
6. Did you find it difficult to use one input method over another (Keyboard, GUI)	2 (1 - 7)	7 (4 - 10)
7. *Did you feel like you had to remember more in your head while under TBFM?*	***3.5 (1 - 5)***	***5 (3 - 7)***
Separation Management (SEPMAN)		
1. Did the system provide indication that you had acknowledged an alert?	4 / 6 - Yes	3 / 6 - Yes
2. *Were you able to understand the symbols?*	***8.5 (5 - 10)***	***7 (3 - 8)***
3. Could you determine the different types of conflict by color?	8.5 (7 - 10)	8 (6 - 10)
4. Did you feel you could quickly determine separation management from the conflict probe?	6.5 (1 - 10)	5 (1 - 7)
5. Were you able to see the indications given to you on the FDB?	7.5 (5 - 10)	6 (1 - 9)
6. Did you find conflict probe Alert view to be excessively long?	6 (1 - 9)	7.5 (5 - 10)
7. Were you able to use your preferred method of input to control the Conflict Probes?	7 (2 - 10)	5.5 (5 - 8)
8. *Did you feel you had to remember more in your head while using the conflict probe?*	***1.5 (1 - 3)***	***4 (2 - 9)***
9. Did you find the conflict probe cluttered your scope?	3.5 (1 - 10)	7 (4 - 10)
10. Did the conflict probe work the way you expected it too?	7 (4 - 10)	5 (2 - 10)
11. Was the Conflict Probe color easily discernable?	7.5 (5 - 10)	8 (5 - 10)
Airborne Reroute (ABRR)		
1. Did the system provide indication that you had acknowledged a reroute?	6 / 6 - Yes	6 / 7 - Yes
2. Were you able to determine the difference between a reroute and airborne reroute indication? 2	9 (6 - 10)	5 (3 - 9)
3. Were you able to tell if an aircraft received a reroute?	8.5 (3 - 10)	6 (3 - 10)
4. Were you able to determine the status of the route based on the color of the indication?	8 (5 - 10)	6 (3 - 9)
5. Were you able to see the indications on the Range Data Block?	7 (5 - 10)	7 (3 - 8)
6. Did you find TFM Quick view to be excessively long?	5 (1 - 10)	6 (3 - 8)
7. Did the airborne reroute clutter your scope?	4.5 (1 - 9)	5 (2 - 10)

We found very few significant differences between the participants in the Low and High Experience groups in their responses to the Exit Questionnaire items. We found significant differences in the responses to two of the questions which pertained to some of the symbology used and the extent to which the participants felt they needed to rely on memory to use certain tools. We conducted a Mann-Whitney analysis and found that the participants in the Low Experience group found the symbology used for TBFM less confusing than the participants in the High Experience group, $U = 24.0$, $p = .011$. The Low Experience group also found they had to remember less information when using this tool than did the High Experience group, $U = 6.5$, $p = .042$. Regarding Separation Management capabilities (Conflict Probe), the participants in the Low Experience group indicated that they were able to understand the symbols better than the High Experience group, $U = 7.0$, $p < .05$, and felt they had to rely on their memory less than the High Experience group, $U = 5.0$, $p = .025$.

We also conducted chi-square analyses to determine whether a significant number of participants in either the Low experience group or the High experience group reported higher than median (> 5.5) or lower than median (< 5.5) ratings for each questionnaire item. We report the results of the questionnaire items for which we found significant differences for each group. Not all of the participants responded to each item.

5.4.3.1 Simulation realism and training

We asked several questions about the simulation realism. Seven of the High Experience participants reported the simulation software to be realistic, X^2 (1, N = 8) = 4.09, $p < .05$). Both the Low Experience and High Experience group reported that there were enough practice scenarios provided to get them familiar with the new ERAM elements, with all eight participants in each group reporting ratings higher than 5.5, X^2 (1, N = 8) = 6.64, $p < .05$. All of the participants in the Low Experience group reported that we provided sufficient training on the DESIREE simulator, X^2 (1, N = 8) = 6.64, $p < .05$. The High Experience group (7 of 8) reported that we provided effective training on Data Comm, X^2 (1, N = 8) = 4.09, $p < .05$.

5.4.3.2 Symbol complexity

For items about display symbology, both groups (eight of eight Low Experience; seven of eight High Experience) reported that the symbology allowed them to determine the difference between Data Comm equipped and non-Data Comm equipped aircraft. All of the Low Experience group participants reported that they were able to determine track ownership of the data blocks, X^2 (1, N = 8) = 6.64, $p < .05$. The Low Experience group (seven of eight) also reported that they found the safety alert easy to see in the data block, X^2 (1, N = 8) = 4.09, $p < .05$, and that they were able to determine the meaning of the different color symbols.

5.4.3.3 Data Comm

For Data Comm, seven of eight participants in the Low Experience group indicated that the symbols were not confusing, X^2 (1, N = 8) = 4.09, $p < .05$. All of the Low Experience participants, X^2 (1, N = 8) = 6.64, $p < .05$, and seven of eight High Experience group participants, X^2 (1, N = 8) = 4.09, $p < .05$, reported that they were able to determine when aircraft were Data Comm equipped. Seven of eight participants in the Low Experience group found the use of color useful in helping them interpret Data Comm symbology, X^2 (1, N = 8) = 4.09, $p < .05$. Seven of eight Low Experience group participants indicated that they were able to send messages quickly with Data Comm, X^2 (1, N = 8) = 4.09, $p < .05$. Both groups (eight of eight Low Experience, X^2 (1, N = 8) = 6.64, $p < .05$; seven of eight High Experience, X^2 (1, N = 8) = 4.09, $p < .05$) reported that Data Comm "flowed" with the rest of the system.

5.4.3.4 TBFM

For TBFM items, five of six of the Low Experience group participants who responded reported that the TBFM symbols, data blocks, and colors were not confusing or distracting, X^2 (1, N = 6) = 5.55, $p < .05$). All six of the participants who responded in the Low Experience group indicated that they did not have to remember more when working with TBFM, X^2 (1, N = 6) = 5.91, $p < .05$). We did not find significant differences in responses for the High Experience group.

5.4.3.5 Separation Management

For items relating to Separation Management, we found that all of the participants in the Low and High Experience groups who responded could determine different types of conflict by color, X^2 (1, N = 6) = 5.91, $p < .05$. All of the Low Experience group participants also reported that they did not have to remember more while using these tools, X^2 (1, N = 6) = 5.91, $p < .05$).

5.4.3.6 ABRR

Finally, for ABRR, we found that all six of the Low Experience group participants who responded reported that they were able to determine the difference between a reroute and an airborne reroute indication, X^2 (1, N = 6) = 5.91, $p < .05$. We did not find a significant result for the High Experience participants.

5.4.4 Debrief Comments

The participants provided additional comments about their experience in the simulation during the debrief session. We summarize their comments below for each of the new features and capabilities. In general, display clutter was noted by some as an issue, but not by all, with some reporting that the data blocks were "not too cluttered," even when there was a lot of information presented. Display clutter came up in discussion most for Conflict Probe and for Path Stretch in TBFM, which we will discuss in the respective sections below. We learned from the Exit Questionnaire that these reports were associated with experience level, and that those in the High Experience group were more likely to report display clutter to be a problem.

The participants reported that working with an RA was important, especially when working busy scenarios and that having an RA increased their use of the tools. Having additional features and functions also available on the RA position allowed the RA to do more, which was reported as helpful. But, some participants reported that they did not want the RA to send uplinks via Data Comm. The participants indicated that, in the operational environment, the increased availability of tools would necessitate additional training to provide a clear designation of roles and responsibilities “to establish mutual understanding and good teamwork between the R-side and RA.” They described that “a paradigm and mind-set shift” would be necessary to achieve this.

The participants reported that they found that entering commands via the trackball was more efficient for some commands, but that having keyboard equivalent entries for these commands is a must. They commented negatively about the inconsistent use of trackball functions (enter, pick) that depended on which function and where in the data block the selection was made. This was confusing and required them to stop and think about which button to choose for the entry they wanted to make. As one comment indicated, “It’s difficult to determine which one is which” on the data block. Using the trackball to make selections through menus was “cumbersome,” especially when they were busy. Many felt it was easier to use available keyboard commands for most of the tasks. As an R-side working alone, it was “difficult to keep up with typing.”

The participants found that the choice of airspace used in the simulation was not conducive to the use of TBFM. The sector using TBFM in this study was fairly small and did not leave much space to absorb the suggested delays or implement the Path Stretch reroutes. The participants suggested using other airspace that would better encourage use of TBFM capabilities for future simulations.

The participants also provided other comments regarding the training provided and other aspects of the simulation. These comments indicated that the participants would have liked to have Navaid labels available on the map to help them better learn the airspace. One participant noted that we did not implement an aircraft type button on the keypad which is something he relies on in the field. The participants found our implementation of the ACL at the RA position unrealistic in that it did not automatically drop or clear aircraft from the list the way they would expect it to in the field. They found that this increased their search time. Finally, the U/C–space key on our keyboard caused confusion for one of the participants, who was used to using this button at his facility as a quick key for the space bar instead of the uplink key.

5.4.4.1 Scenarios and Team Configuration

The participants reported that they used the tools less when they were busy. They reverted to "talking and turning" when busier and used Data Comm less. However, they also reported that they did increase their use of the tools over time as they continued to gain familiarity with them.

Having an RA was helpful for the busier scenarios and also helped the participants increase use of the tools. Sector 10 had more time for strategic use of available tools because it was not as demanding as Sector 27. One of the participants specifically commented on losing situation awareness when traffic was busy and he was working alone in Sector 27.

The participants commented that having an RA was especially helpful when planning and rerouting. However, they also felt that the responsibilities of the RA would change as a result of the increased tool availability necessitating a clear designation of roles and responsibilities for that position. There was general agreement that the RAs would need to have previous R-side experience to appropriately help manage the sector. The participants indicated that they felt that an RA without R-side training would take handoffs too quickly. We found that in the objective data, the R-side participant almost always took handoffs and only had the RA do so in a small percentage of situations. Overall, the participants commented that the additional RA position responsibilities are a "training and culture issue" more than a hardware or software issue.

5.4.4.2 43" Display

Two participants in the simulation had experience with the 43" display at their facility. Both reported experiencing negative effects (e.g., headache, nausea) when working with those displays at the facility, but reported that the problems were not as apparent in the laboratory environment. One of the participants reported mild headaches during the simulation. He found that the headaches were worse when he was working traffic in ER3 where the monitors were 6" closer than they were in ER2 (due to desktop space). Two of the participants reported mild eye strain when using the displays in ER2 where the displays were 6" further away.

There are many factors that could cause differences between the negative effects experienced in the field compared to the lab, including that the drivers and configuration for the displays differ between the two implementations. These problems have been reported in the operational environment and efforts are underway to determine how to remediate them. Beyond these issues, the participants reported positively about the display, including that the larger size offered great flexibility for positioning menus and toolbars, and that having frequencies available at the edges of the display (e.g., the "horse collar") was useful. Several participants also reported that having the checklists in the horse collar boundaries made the increased screen size less apparent.

Some negative comments about the displays included that it was problematic to use macros because things were "too far away" so it took more time to access them on the larger displays. The participants also commented that targets were small and that it could be difficult to select them with the trackball, that you had to be "right on" a specific area to do so. This caused them to use the keyboard more frequently. Other comments indicated that there was too much presented in the horse collar and that they would prefer that this area be customizable so they could include information as needed.

5.4.4.3 Conflict Probe on R-side display

The participants had mixed reactions to Conflict Probe on the R-side display. Although positive about having an indication of potential conflicts on the R-side, a few participants reported that the information caused clutter on the display and that it was often difficult to tell which aircraft pairs were in conflict based on the proximity graphics only. They also felt that it was problematic that the route line designations did not time out and that the graphics had to be removed manually by picking on the indicator again either through the conflict probe alert view or the indicator on the FDB. However, some participants did not find clutter to be as much of a problem. This likely reflects the different responses we obtained on the Exit Questionnaire between the participants in the High Experience group who reported that new display elements caused the display to be too cluttered and the Low Experience group who did not report that to be as much of a problem.

5.4.4.4 Data Communications and Voice Communication Indicator

The participants were generally positive about Data Comm, but found it more difficult to use in busy situations. They found it useful that Data Comm was available from the RA position and that it was especially useful for issuing reroutes. One of the participants commented that the ability to uplink reroutes via Data Comm was a "game-changer." Other comments indicated that the participants would also find the ability to uplink a route they created from selected latitudes/longitudes very helpful.

The participants' negative comments about Data Comm indicated that the use of menus for these functions was not ideal. Even though Data Comm keyboard entries were available for most functions, they were not obvious and required a high learning curve. The participants also discussed wanting to be able to uplink headings, a function not currently planned for implementation. The participants also reported the they did not find the white Data Comm session symbol informative.

The participants made several negative comments regarding the VCI, including that the VCI symbol was difficult to see, that it was too small, and that the white color blended too much with the rest of the data block. Another controller reported that silent check-in of aircraft with Data Comm was difficult to notice. Several participants indicated that all aircraft should be required to verbally check in. The participants noted that they often waited for silent monitoring for intrafacility handoffs (e.g., aircraft descending to IAD) and suggested that only aircraft in level flight be eligible for monitoring. Any aircraft not in level flight should be required to verbally check in, or an auditory cue should be provided to alert the controller that the aircraft is monitoring. The participants found the auto check-in symbol difficult to notice. They suggested implementing a brighter green or providing an audible alert to make it more noticeable. (It was noted during the discussion that the symbol has already been changed from white to green in the field.)

5.4.4.5 Time-Based Flow Management

The participants reported that the "C" notification in line zero, indicating a speed advisory, was sometimes confusing and that it could be difficult to differentiate from the "C" indicating conflict alert. They commented that the GIM-S speed advisories were useful, but that they were often too high, sometimes "too high to be realistic," and that the commands to adjust the delay times were confusing. When there was sufficient time to implement the suggested delay, the participants reported that the tool worked well. Overall, the participants found having the GIM-S speed advisories useful, although not always reasonable to implement. Only one of the participants had prior experience with GIM-S at his facility.

The participants found the Path Stretch feature complicated, not intuitive to use, and that it gave "bad advice" by sending aircraft through restricted airspace. The route lines also added too much display clutter. As we reported previously, the Path Stretch implementation for the simulation did not consider the SAA in its calculations and, therefore, sometimes suggested alternate routes through the SAA.

Despite the SAA issues, the participants indicated that the Path Stretch concept would be useful if full calculations were taken into account during implementation. However, at the time of the simulation, the capability was more problematic than helpful. A few participants indicated that it was simply faster to vector aircraft than to use the tool. They further suggested that the tool should be site-specific and adaptable because the practicality or usefulness of the tool would depend on the specifics of the airspace. In our simulation, they reported that Sector 27 was too small and too close to the aircraft destinations to make it helpful. It would be more useful to have

Path Stretch capabilities in place in a larger, higher-altitude sector. In fact, these capabilities are planned to be used in the field for metering up to top-of-descent (TOD).

It is important to note that the participants stated that they did not think that TBFM capabilities would be used without Data Comm. They also indicated that they preferred to use keyboard commands for many of the functions because there is “too much to click on.” One participant suggested that it would be helpful to set up parameters based on the airspace so that when implementing a Path Stretch maneuver, the aircraft could cross sectors and not result in the need for extraneous point outs.

5.4.4.6 Airborne Reroute

Overall, the participants found the ABRR capability very useful. However, there was some confusion about the inconsistent use of color symbology (cyan or white) in the menus and FDB. The participants suggested that it should be consistent across all menus. Currently, the color differs based on the view. One comment indicated that selecting the cyan T should open a route menu in addition to showing the route probe, and others commented on confusion between when to left or right click.

The participants indicated that the TFM reroute menus were very helpful, especially (and for a few participants, only) when Data Comm was available. One controller said he “wanted Data Comm and ABRR at his facility now.” But, others commented that they would want to be able to issue a “descend via” by Data Comm. Using procedures such as “descend via” or “climb via” can be confusing when issued by voice, and they are sometimes used inconsistently. However, this function is not planned for full Data Comm implementation. The participants reported that they ignored most reroutes when the scenarios got busier.

5.4.4.7 Controller-to-Controller Communication

The participants had mixed reactions to the CTCC capability. A couple of the participants questioned the usefulness and need for the tool, whereas others indicated that they thought the function was “great.” A couple of comments that captured the reaction of most participants was that CTCC is “a good idea, but the implementation is poor,” and that, “getting control outside sector is great, but I’d rather just call” because making a phone call is faster.

The participants found CTCC less helpful than it might have been primarily because of confusing design attributes. Some participants commented that the symbology was difficult to notice, particularly the CTCC request symbol (small green triangle). Some participants also noted the tool was “cumbersome” and “not intuitive” because it was not clear how to enter commands or how to enter data into the menus. For example, one participant described having difficulty activating the “Release Control” function due to uncertainty about what to enter in the pop-up textbox. In general, participants commented negatively on the textboxes in CTCC. They did not find it useful to have them prepopulated with information from the previous sector. They reported that they would rather have the fields appear blank because it required more effort and workload to have to eliminate the pre-populated information and then enter the information needed.

Others commented that there was a sense of “tunnel vision” when using the CTCC menus which affected their radar scan. The menus took too much interaction time. One participant suggested adding the CID to the information—“Show who sent what so I don’t have to do all menus, so I can use keyboard.”

Finally, a few other comments expressed confusion as to why CTCC would be needed in addition to an automatic point out because, “the only new function is the ability for the silent coordination if there are choices or reference aircraft.” However, it should be noted that automatic point outs do not provide requests for approval or conditional point-out approval requests.

6 Conclusions

This phase of the OAIP project investigated the effects of co-locating new tools and capabilities onto the existing en route air traffic control system. This included Conflict Probe on the R-side display, ABRR, TBFM, CTCC, Data Comm, and a 43-inch monitor at the R-side position. Software engineers at the RDHFL simulated the tools or connected to available operational capabilities and added them to the DESIREE en route simulator to provide a realistic evaluation environment. We conducted a baseline simulation with 16 current, en route CPCs who managed traffic in this environment in moderate and high traffic level scenarios, either as R-sides alone or in R-side/RA teams. Eight of the participants had 5 or fewer years of experience as en route CPCs and eight of them had 15 or more years of experience, allowing us to assess differences in performance and subjective feedback based on experience level.

We found that the results for the basic air traffic measures were consistent with what we expected given the scenario traffic levels. The participants managed more aircraft and made more voice transmissions in the busier traffic scenarios than the moderate traffic level scenarios. The participants tended to accept handoffs into the sector later and hand off aircraft out of the sector sooner when they worked the higher traffic level scenarios and when they worked as R-sides alone. These results suggested that the participants were looking to manage workload by regulating the traffic in their airspace. Subjective workload reports also indicated that the participants experienced higher workload when they worked the higher traffic level scenarios and when they worked as R-sides alone. We also found more procedural deviations in these busy conditions.

We analyzed eye movement and EEG data to investigate other aspects of workload and cognitive demand. We hypothesized that higher demand may result when the participants process additional data block information provided by the new tools. We correlated pupil diameter with FDB complexity to determine whether we would find evidence of higher workload (i.e., larger pupil diameter) when the participants fixated data blocks with a greater number of display elements. We did not find a significant correlation between those two measures. However, it is important to bear in mind that pupil diameter is affected by multiple factors beyond processing the visual stimuli. The additional mental processing that the participants were engaged in also influenced pupil diameter. Air traffic controllers are continuously gathering and assimilating information and planning actions when managing the airspace. A participant may have been fixating a low-complexity FDB to gather information, but also assimilating that with other information to develop a solution to a complex situation. In their review of pupilometry and cognitive control tasks, van der Wel and van Steenbergen (2018) indicate that "it is impossible to decide whether physiological signals reflect mere 'task demand' or 'effort exertion'" (p. 2006). In our simulation, we are not able separate "task demand," as might be suggested objectively by FDB complexity, from "effort exertion" that occurred during mental processing.

When we looked at pupil diameter on a more global level, we did find differences between test conditions. We found larger average pupil diameters when the participants managed the high traffic level scenarios compared to the moderate traffic level scenarios. Pupil diameter reflected the greater demands of the busier traffic scenarios. This result aligns with findings obtained in other air traffic studies such as those conducted by Ahlstrom and Friedman-Berg (2006). They found that pupil diameter reflected levels of workload across different traffic level scenarios as well as between conditions that implemented different types of weather displays.

With regard to the EEG analyses, we began with the goal of examining participant brain activity when they initially fixated the FDBs. After collecting the EEG and eye-tracking data, we created fERPs time-locked to participant fixations on FDBs. We used an ICA-based approach to localize the fERP activity to potential neural generators. Our analysis revealed several brain regions that likely yielded the EEG activity we observed; Right Posterior Cingulate Gyrus, Left Anterior Temporal Pole, Right Angular Gyrus, and Right Primary Motor Cortex. These brain regions are involved in a variety of cognitive processes such as: attention, decision-making, change detection, semantic meaning extraction, object processing, and response planning (see Table 4 for details).

We conducted two analyses on the EEG data, one looking at the effects of our experimental conditions and the other to see how fERPs change with FDB complexity. We did not find any evidence that our fERPs were affected by FDB complexity. Similar to the pupil diameter data, the fERPs are likely reflecting activity beyond the processing of the specific FDB, so it is not too surprising that we did not find a direct relationship between the EEG and FDB complexity. While we observed some differences in the fERP amplitudes at certain time windows, our conclusions are only weakly supported by the data. Due to the large volume of data and an absence of a priori predictions, our *p* values are inflated and may represent noise. An inherent problem with large data sets is that they can produce spurious results—those that are statistically significant but that may not be of practical significance. EEG data alone do not give us much detail about the brain's involvement during the task. EEG studies rely upon comparisons across conditions and correlations with behavior to reach conclusions about neural processing. However, this study utilized novel data acquisition protocols in a noisy and chaotic environment. Electrical activity from laboratory equipment as well as the muscles used for eye movements and speaking can interfere with the EEG signal. Typical EEG laboratories are located inside electrically shielded spaces and participants are instructed to minimize speaking and moving their eyes to reduce signal noise. Since we were simulating a high-fidelity air traffic control environment during our EEG recordings, we acquired the data under comparatively busy and electrically noisy conditions that required a novel analysis strategy.

Additional analyses of our data may be useful to further investigate the relationship between fERP activity and FDB complexity. For example, it may be useful to analyze the fERP data with respect to the first fixation on a data block during a scenario. That initial fixation may involve additional processing of the FDB as the controller is obtaining information about an aircraft for the first time. That situation may better reflect a response to the number of elements presented in the FDB. Other analyses may also be useful to examine the effects of time-on-task, workload, or fatigue which we did not evaluate in this simulation. These factors have been investigated in other ATC studies but were conducted in EEG laboratories using low-fidelity ATC tasks (Dasari, Crowe, Ling, Zhu, & Ding, 2010; Dasari, Shou, & Ding, 2017). It would be useful to determine whether the data we collected in our "noisy" environment showed similar results.

Although we did not find the relationship between fERPs and FDB complexity measures that we expected in our simulation, we did identify fERPs with reasonable amplitude and timing patterns from brain regions that one might expect to be involved in a complex task such as ATC. This study and the EEG results can serve as a baseline to compare results from any additional studies and to and develop a priori predictions for future fERP results from HITL simulations.

The primary focus of this simulation involved understanding participant use of the new tools and capabilities and whether there were difficulties associated with using them. We evaluated the use of the new tools and capabilities by examining the number of times the participants interacted with them during the simulation. An interaction occurred when a participant made objective use of the tool, for example, using the trackball to select a display element or making a keyboard entry to provide data to the system. Although the participants could have made use of data provided by the tools—such as information provided by Conflict Probe in the ACL—to guide decisions about maneuvers, we would not have been able to directly observe such usage. Therefore, we implemented a definition of tool use that provided a measurable metric. We found that overall tool use for Conflict Probe, CTCC, ABRR, TBFM varied by test condition. The R-side participants interacted more with the new tools when they worked alone than when they worked with an RA, though the R/RA teams used the new tools more overall. Regarding use of the individual tools, we found that participants did not often use Conflict Probe on the R-side. Half of the participants did not interact with the tool at all. For TBFM, the majority of the interactions (70%) were to reject system recommendations. For Data Comm, we found that the participants interacted more with the tool when they worked as R-sides alone and when they worked moderate traffic level scenarios. For ABRR and CTCC, the RA participants interacted more with the tools than the R-sides when the participants worked in teams.

The participants provided subjective feedback on their use of the tools on the questionnaires and in the debrief session to help us understand their reasons for using or not using the tools. For Conflict Probe, the participants indicated that they often found the route displays caused clutter, especially if multiple aircraft were involved. As a result, they did not use this tool often. However, it is still possible that the participants made use of Conflict Probe data, as discussed above, in ways that were not directly observable.

The participants commented favorably on the TBFM concept, but they reported that the tool did not work well in this simulation because the sector in which it was used was fairly small, too close to the destination airports, and included SAA. The delay times provided by TBFM were often too high for the aircraft to absorb, and the Path Stretch advisories did not account for the SAA, often suggesting a reroute through that area. The participants did not feel they could confidently accept most recommendations as a result. The participants commented that TBFM would be better implemented in a larger sector, further from aircraft destinations that could more effectively absorb delays such as Cleveland Center, ZOB 73, in our simulation. The participants also found that, at a glance, the “C” designation in line zero of the data block could be confused with the conflict alert (“CA”) designation. The data from the questionnaire ratings indicated that this may have been more of an issue for the participants in the High Experience group who reported that symbol complexity was more confusing and found TBFM information less useful than the Low Experience participants.

The participants reacted favorably to Data Comm, especially for uplinking routes. But, they also commented on a lack of “obvious keyboard commands” for some functions and the need to rely too much on menus to perform tasks. There are a number of reasons that could have driven the desire for keyboard command entries, including that menus occupy space on the display, require focus on the information presented, and take time to interact with and navigate. Focusing on menus can disrupt the controllers’ scan and divert attention from the traffic. Keyboard entries, on the other hand, allow controllers to continue to scan the display while they complete the required tasks. Keyboard entries are also typically executed more quickly than menu interactions, a result we observed in our simulation. Menus are often considered more useful for less-experienced system users because they provide task guidance, whereas keyboard entries are considered more useful for experienced users who have more domain knowledge and more familiarity with command structure. However, menu design plays a major role in their usefulness and how quickly needed information can be accessed. These factors also affect which method system users will opt to work with if given a choice (Paap & Roske-Hofstrand, 1988). Although we did not find a significant difference in the use of menu and keyboard entries between the Low and High Experience groups in our simulation, we recommend providing both methods when possible so that controllers can determine which is more effective for a particular task at a given time.

With respect to Data Comm, we found that the R-side participants used Data Comm less in the higher traffic level scenarios, likely due to a need to ensure more timely clearances when scenarios were busier and they could not wait for Data Comm transmission delays. The participants commented that when they were busy, they resorted to “talking and turning.”

The participants generally found ABRR helpful as long as it could be used with a Data Comm uplink. The tool allowed for common reroutes to be assigned quickly without having to call the aircraft to confirm. However, the participants expressed confusion about the different color designations (cyan and white) used for this tool. They also noted that ABRR would need to be facility dependent, as there are different rules for “descend-vias” across the country which would need consideration, although these are not currently slated for Data Comm rollout.

The participants had mixed reactions to CTCC. Some of the participants felt the tool had a lot of potential but that it was poorly executed. Comments indicated that the symbology—green triangle—was not salient enough to be noticed. Several participants felt that a coordination call would be faster because it required time to navigate menus and clear autofill information that the system pre-populated based on the location of the aircraft. This caused perceived delays and resulted in considerable “heads down” time, taking away from the traffic scan. We also saw evidence of difficulties using CTCC due to confusion about which trackball button (Pick or Enter) to select when interacting with this tool. This confusion would become less problematic over time as controllers gained more familiarity with the tool. However, effective interface design can help enable tools to be used more efficiently from the start. One such implementation would be to make the tool available in an area of the data block that is meaningfully associated with the tool’s function. CTCC is not meaningfully associated with the current access point, the speed field, so it is not an obvious location for it. Another implementation would be to provide a consistent interaction method for similar functions to establish “rules” for which button (Pick or Enter) to use. For example, the rule for the Pick button might be that it is always used to access functions that are located “behind” an access point—to get to other functions that may not be immediately associated with that data block location—as is the case for the implementation of CTCC used in our simulation.

In addition to the use of the specific tools, the participants commented on the need to clearly designate the roles and responsibilities of the R-side and RA controllers to ensure that they work effectively as a team. Each team member must be aware of what the other is doing to ensure that the traffic is handled safely and efficiently, especially as new tools and capabilities enable the RA to perform more tasks. Display indicators can be useful in this regard as they can designate when an action has been completed. Although we did not incorporate such indicators in our simulation, Willems and Hah (2008) did. In their simulation investigating the integration of new tools in the en route domain, Willems and Hah included symbology that designated when the RA controller accepted a handoff (i.e., a dashed box around the FDB) so that the R-side would have confirmation that the action had been taken. Such feedback would help support effective and efficient teamwork between the R-side and RA controllers.

The participants in our simulation commented that they were readily able to trust the other member of their team, but that is not necessarily the case in the operational environment. Even in our simulation, we observed that the R-side participants were likely to allow the RA to hand off aircraft to other sectors, but were much more likely to take handoffs into the sector themselves. With respect to the new tools and capabilities, the interactions were mixed and the RA interacted with these new features fairly frequently when the participants worked in teams. A high level of teamwork and coordination is needed, and the R-side controller must fully understand and agree upon the actions the RA will take and the responsibilities they will shoulder. This is important in the operational environment for obvious safety reasons. It is also important because, in the field, the controller teams will often have different levels of experience, which may cause them to respond to or use the tools differently from one another, causing them to work less effectively as a team. As Morgan, Herschler, Wiener, and Salas (1993) note in their summary of the effects of automation on aircrew coordination, "crew members must continue to be trained in proactive cockpit management and to effectively coordinate their use of aircraft automation" (p.108). Without doing so, the integration of automation can result in negative outcomes. It is likewise critical to ensure that the R-side and RA controllers are trained both on the automation and on how best to use it when working as part of a team.

With respect to participant experience level, many of the differences we found pertained to the participants' subjective impressions of the new tools. The participants in the Low Experience group generally reported that the information provided had a more positive effect on their performance, control of traffic, situation monitoring, managing sector resources, and rerouting and evaluating flight plans than did the participants in the High experience group. However, the participants in the Low Experience group also reported greater effort and frustration, and greater physical and temporal demand than did the participants in the High Experience group. The participants in the High Experience group also rated some of the symbology to be more confusing than the participants in the Low Experience group and that some functions required more reliance on working memory. These differences may influence the extent to which individuals make use of the tools and could negatively affect R-side/RA team coordination.

Finally, this simulation confirmed several important human factors issues initially identified by Willems and Dworksy (2018) in Phase 1 of the OAIP project. These issues included confusing symbology (e.g., VCI), display clutter (e.g., Conflict Probe), confusing use of color (e.g., ABRR), and confusing command entries (e.g., CTCC). These issues should be addressed to improve the usefulness of the new tools and their integration into the existing air traffic system.

7 References

Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics, 36*(7), 623-636. doi:doi:https://doi.org/10.1016/j.ergon.2006.04.002

Arsalidou, M., & Taylor, M. J. (2011). Is 2+ 2= 4? Meta-analyses of brain areas needed for numbers and calculations. *Neuroimage, 54*(3), 2382-2393. doi:10.1016/j.neuroimage.2010.10.009

Bonner, M. F., & Price, A. R. (2013). Where is the anterior temporal lobe and what does it do? *Journal of Neuroscience, 33*(10), 4213-4215. doi:https://doi.org/10.1523/JNEUROSCI.0041-13.2013

Chatrian, G. E., Lettich, E., & Nelson, P. L. (1988). Modified nomenclature for the “10%” electrode system. *Journal of Clinical Neurophysiology, 5*(2), 183-186. doi:10.1007/s10548-014-0400-8.

Christensen, H. (2015). *Analysis of ordinal data with cumulative link models—estimation with the R-package ordinal.* Retrieved from cran.r-project.org/web/packages/ordinal/vignettes/clm_intro.pdf

Crutchfield, J. M., & Millan, A. (2017). Proposed Scenarios for the Standardization of the Evaluation of New ATC Technologies and Procedures. *19th International Symposium on Aviation Psychology,* (pp. 83-88). Retrieved from https://corescholar.libraries.wright.edu/isap_2017/81

Dasari, D., Crowe, C., Ling, C., Zhu, M., & Ding, L. (2010). EEG Pattern Analysis for Physiological Indicators of Mental Fatigue in Simulated Air Traffic Control Tasks. In . (Ed.), *PROCEEDINGS of the HUMAN FACTORS and ERGONOMICS SOCIETY 54th ANNUAL MEETING* (pp. 1-5). Human Factors and Ergonomics Society, Inc. doi:10.1518/107118110X12829369200477

Dasari, D., Shou, G., & Ding, L. (2017). ICA-Derived EEG Correlates to Mental Fatigue, Effort, and Workload in a Realistically Simulated Air Traffic Control Task. *Frontiers in Neuroscience, 11*(297), 1-12. doi:10.3389/fnins.2017.00297

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9-21. doi: 10.1016/j.jneumeth.2003.10.009

Faraway, J. J. (2016). *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models.* Boca Raton, FL: Chapman & Hall/CRC Press.

Federal Aviation Administration. (2019, August 15). *Air Traffic Organization Policy Order JO 7110.65Y.* Retrieved from https://www.faa.gov/documentLibrary/media/Order/7110.65Y_ATC_Bsc_w_Chg_1_2_dtd_7-16-20.pdf

Mangiafico, S. S. (2016). *Summary and Analysis of Extension Program Evaluation in R, version 1.18.1.* Retrieved from https://rcompanion.org/documents/RHandbookProgramEvaluation.

Mazoyer, B., Zago, L., Mellet, E., Bricogne, S., Etard, O., Houde, O., & Rzourio-Mazoyer, N. (2001). Cortical networks for working memory and executive functions sustain the conscious resting state in man. *Brain research bulletin, 3,* 287-298. doi:https://doi.org/10.1016/S0361-9230(00)00437-8

Meullbacher, W., Ziemann, U., Wissel, J., Dang, N., Koflet, M., Facchini, S., & Hallett, M. (2002). Early consolidation in human primary motor cortex. *Nature, 415*(6872), 640-644. doi:10.1038/nature712

Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology, 48*(2), 229-240. doi:10.1111/j.1469-8986.2010.01061.

Morgan, B. B., Herschler, D. A., Wiener, E. L., & Salas, E. (1993). Implications of automation technology for aircrew coordination and performance. In W. B. Rouse (Ed.), *Human/technology interaction in complex systems* (Vol. 6, pp. 105-136). Greenwich, CT: JAI.

Paap, K. R., & Roske-Hofstrand, R. J. (1988). Design of Menus. In M. Helander (Ed.), *Handbook of Human-Computer Interaction* (pp. 205-233). New York: Elsevier.

Pearson, J. M., Heilbronner, S. R., Barack, D. L., Hayden, B. Y., & Platt, M. L. (2011). Posterior cingulate cortex: adapting behavior to a changing world. *Trends in cognitive sciences, 15*(4), 143-151. doi:10.1016/j.tics.2011.02.002

Peelen, M. V., & Caramazza, A. (2012). Conceptual object representations in human anterior temporal cortex. *Journal of Neuroscience, 32*(45), 15728-15736. doi:https://doi.org/10.1523/JNEUROSCI.1953-12.2012

Penfield, W., & Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation . *Brain, 60*(4), 389-443. doi:https://doi.org/10.1093/brain/60.4.389

Raichle, M. E. (2015). The brain's default mode network. *Annual review of neuroscience, 38*, 433-447. doi:https://doi.org/10.1146/annurev-neuro-071013-014030

Seghier, M. L. (2013). The angular gyrus: multiple functions and multiple subdivisions. *The Neuroscientist*, *19*(1), 43-61. doi:10.1177/1073858412440596

van der Wel, P., & van Steenbergen, H. (2018, February 18). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review, 25*, pp. 2005–2015. doi:https://doi.org/10.3758/s13423-018-1432-y

Willems, B., & Dworsky, M. (2018). Optimization of Air Traffic Control Information Presentation Phase 1. *Unpublished manuscript*.

Willems, B., & Hah, S. (2008). *Future En Route Workstation Study (FEWS II): Part I - Automation Integration Research.* (DOT/FAA/TC-08/14,II). Federal Aviation Administration. Atlantic City International Airport: US Department of Transportation.

A Informed Consent Statement

I, ______________________________, understand that this controller simulation, entitled "Air Traffic Control – Simulations for Presentation and Information Display Optimization (ATC-SPIDO)" is sponsored by the Federal Aviation Administration (FAA), AJM-1320. The sponsor, Ben Willems, can be reached at his office (609-485-4191), cell phone (609-369-1660), or via e-mail at Ben.Willems@faa.gov. The Principal Investigator for the project is Carolina Zingale, Human Factors Branch, ANG-E25. She can be reached at 609-485-8629 or by email at carolina.zingale@faa.gov.

Nature and Purpose

I volunteer as a participant in this en route Air Traffic Control simulation experiment. The primary purpose of this simulation experiment is to provide an objective assessment of the impact on en route air traffic controllers of the integration of current and new capabilities into the En Route Automation Modernization (ERAM) system. In addition to assessing the impact on en route controllers, this experiment will also assess how the integration of capabilities in ERAM may affect en route controllers differently depending on their level of experience and sector team composition. Finally, this experiment will attempt to validate the HF issues in the ERAM user interface and functions that the research team uncovered during an earlier phase of the project.

Simulation Procedures

I will participate as a volunteer and control traffic in a sector either as part of a Radar and Radar Associate sector team or by myself. My time commitment is two weeks. As a participant, I will travel in on Monday, participate in the experiment from Tuesday until Thursday of the second week, and travel back on Friday of the second week. I will work from about 8:00 AM to about 4:30 PM with a lunch break and at least two rest breaks. During the introduction, I will get the opportunity to review project objectives, participant rights and responsibilities. After the introduction, I will begin controlling aircraft on a practice scenario (< 30-minutes). While controlling traffic on the Radar position, a Smarteye Pro eye tracker will record eye movements. After I complete training, I will control traffic in several experimental simulation scenarios. During the experimental scenarios, automated systems will collect data on my interactions with the simulator as well as operational system variables. The simulation environment will also record audio and video. During the last training scenario and the experimental scenarios, electroencephalography (EEG) equipment and software will record the electrical activity generated by my brain while controlling traffic. While I am controlling traffic, subject matter experts may conduct over-the-shoulder ratings.

Anonymity and Confidentiality

My participation is strictly confidential. Any information I provide will remain anonymous: data or reports will not associate individual names or identities.

All records will be maintained with complete confidentiality. Your name will not be associated with any of the information collected. A random number will be assigned to your data. Your name will never be associated with that number.

The information that you provide as a participant is strictly confidential and you shall remain anonymous. No Personally Identifiable Information [PII] will be disclosed or released, except as may be required by statute. Any Personally Identifiable Information [PII] will be protected according to FAA Order 1370.121 – FAA Information Security and Privacy Program & Policy. You will not be identifiable by name or description in any reports or publications about this study. You may withdraw from this study at any time without penalty. Data provided until the point of termination will be stored and could potentially be used in the analysis. If you determine that you do not want your data used, you may inform the researcher and your data will not be used.

Benefits

I understand that the only benefit to me is that I will be able to provide the researchers with valuable feedback and insight into the effects of integration of multiple display features and functions. My data will help the FAA to determine how to resolve human factors issue resulting from integration of information and functions from multiple programs into the ERAM Computer Human Interface.

Participant Requirements and Responsibilities

I am aware that to participate in this study I must be either a novice controller with only 3-5 years of experience or a seasoned controller with 15-20 years of experience. I will control my aircraft and answer questions asked during the study to the best of my abilities. I will not discuss the content of the simulation with other potential participants until the completion of the study.

Participant Assurances

I understand that my participation in this study is voluntary and I can withdraw at any time without penalty. I also understand that the researchers in this study may terminate my participation if they believe this to be in my best interest. I understand that if new findings develop during the course of this research that may relate to my decision to continue participation, the researchers will inform me. I have not given up any of my legal rights or released any individual or institution from liability for negligence.

The research team has adequately answered all the questions I have asked about this study, my participation, and the procedures involved. I understand that Carolina Zingale or another member of the research team will be available to answer any questions concerning procedures throughout this study. If I have questions about this study or need to report any adverse effects from the research procedures, I will contact Carolina Zingale at (609) 485-8629.

Discomfort and Risks

The device that monitors eye movements uses near-infrared light. The intensity of the infrared illumination is about one thirtieth of the intensity expected while walking outside on a sunny day and should not cause any discomfort or risk to my health.

The EEG system uses a conductive gel to ensure contact between the scalp and the electrodes. Application of the conductive gel may cause some redness at the application site. The experimenters will provide shampoo and towels to remove any remaining gel from my hair.

I agree to report any injury or suspected adverse effect to Carolina Zingale immediately at (609) 485-8629. Local clinics and hospitals will provide any treatment, if necessary. I agree to provide, if requested, copies of all insurance and medical records arising from any such care for injuries/medical problems.

Signature Lines

I have read this informed consent form. I understand its contents, and I freely consent to participate in this study under the conditions described. I understand that, if I want to, I may have a copy of this form.

Research Participant:____ Date:__________

Investigator:____ Date:__________

Witness:____ Date:__________

B Background Questionnaire

This questionnaire is designed to obtain information about your background and experience as a certified professional controller (CPC). Researchers will only use this information to describe the participants in this study as a group. Your identity will remain anonymous.

1. What is your **gender**?	○ Male		○ Female
2. What is your **age**?	_____ **years**		
3. How long have you worked as **an Air Traffic Controller (include both FAA and military experience)**?	_____ **years** _____ **months**		
4. How long have you worked as a **CPC for the FAA**?	_____ **years** _____ **months**		
5. How long have you **actively controlled traffic** in the **en route** environment?	_____ **years** _____ **months**		
6. How long have you **actively controlled traffic** in the **terminal** environment?	_____ **years** _____ **months**		
7. How many of the **past 12 months** have you actively controlled traffic?	_____ **months**		
8. How long have you been **using ERAM operationally**?	_____ **years** _____ **months**		
9. When did you last receive **ERAM training**?	_____ / ________ **(month/year)**		
10. Rate your current **skill as a CPC**.	Not Skilled	①②③④⑤⑥⑦⑧⑨⑩	Extremely Skilled
11. Rate your **level of motivation** to participate in this study.	Not Motivated	①②③④⑤⑥⑦⑧⑨⑩	Extremely Motivated

C Post-Scenario Questionnaire

Overall Performance, Workload, Situation Awareness, and Simulation Ratings

1. Overall, how **easy was it to use available information** during this scenario run**?**	Extremely Easy	①②③④⑤⑥⑦⑧⑨⑩	Extremely Difficult
2. How **easy was it to view presented information** during this scenario run?	Extremely Easy	①②③④⑤⑥⑦⑧⑨⑩	Extremely Difficult
3. How often did the **presented information cause clutter on the display** during this scenario run?	Never	①②③④⑤⑥⑦⑧⑨⑩	Always
4. What effect did **the presented information have on your ability to control traffic safely** during this scenario run?	Negative Impact	①②③④⑤⑥⑦⑧⑨⑩	Positive Impact
5. Was the Radar Associate (RA) side being staffed essential to handle the traffic in this scenario?	Not Necessary	①②③④⑤⑥⑦⑧⑨⑩	Necessary
6. Rate your **level of Air/Ground communication**.	Extremely Low	①②③④⑤⑥⑦⑧⑨⑩	Extremely High
7. Rate your **level of situational awareness** during this scenario.	Extremely Poor	①②③④⑤⑥⑦⑧⑨⑩	Extremely Good
8. Rate the **performance of the simulation pilots** in terms of their responding to control instructions and providing readbacks.	Extremely Poor	①②③④⑤⑥⑦⑧⑨⑩	Extremely Good
9. Rate the **difficulty** of this scenario.	Extremely Easy	①②③④⑤⑥⑦⑧⑨⑩	Extremely Difficult
10. How did the **amount of presented information** affect your ability to perform the following tasks?			
a) Situation monitoring	Negative Impact	①②③④⑤⑥⑦⑧⑨⑩	Positive Impact
b) Searching for potential aircraft conflicts	Negative Impact	①②③④⑤⑥⑦⑧⑨⑩	Positive Impact
c) Resolving potential conflicts	Negative Impact	①②③④⑤⑥⑦⑧⑨⑩	Positive Impact

d) Managing air traffic sequences	Negative Impact	①②③④⑤⑥⑦⑧⑨⑩	Positive Impact
e) Re-routing or evaluating flight plans	Negative Impact	①②③④⑤⑥⑦⑧⑨⑩	Positive Impact
f) Managing sector/position resources	Negative Impact	①②③④⑤⑥⑦⑧⑨⑩	Positive Impact
11. Rate your **mental demand** during this scenario (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.).	Extremely Low	①②③④⑤⑥⑦⑧⑨⑩	Extremely High
12. Rate your **physical demand** during this scenario (e.g., communications and key presses).	Extremely Low	①②③④⑤⑥⑦⑧⑨⑩	Extremely High
13. Rate your **temporal demand** during this scenario. (How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred?)	Extremely Low	①②③④⑤⑥⑦⑧⑨⑩	Extremely High
14. Rate your **performance** during this scenario. (How successful do you think you were in accomplishing the goals of the task?)	Extremely Low	①②③④⑤⑥⑦⑧⑨⑩	Extremely High
15. Rate your **effort** during this scenario. [How hard did you have to work (mentally and physically) to accomplish this level of performance?]	Extremely Low	①②③④⑤⑥⑦⑧⑨⑩	Extremely High
16. Rate your **frustration level** during this scenario. (How insecure, discouraged, irritated, stressed and annoyed did you feel during the task?)	Extremely Low	①②③④⑤⑥⑦⑧⑨⑩	Extremely High

Additional comments:

D Exit Questionnaire

Please respond to each of the following items based upon your overall experience during the simulations. Fill in one circle to indicate your response to each item.

Part 1- Simulation Realism and Research Apparatus Ratings

1. Rate the **overall** realism of the simulation.	Not at all Realistic	①②③④⑤⑥⑦⑧⑨⑩	Extremely Realistic
2. Rate the **realism of the simulation hardware** compared to actual equipment.	Not at all Realistic	①②③④⑤⑥⑦⑧⑨⑩	Extremely Realistic
3. Rate the **realism of the simulation software** compared to actual equipment.	Not at all Realistic	①②③④⑤⑥⑦⑧⑨⑩	Extremely Realistic
4. Rate the **realism of the simulation traffic scenarios** compared to actual NAS traffic.	Not at all Realistic	①②③④⑤⑥⑦⑧⑨⑩	Extremely Realistic
5. To what extent did the **WAK workload rating technique interfere** with your ATC performance?	Not At All	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
6. Did you feel that there were **enough practice scenarios to familiarize you** with the new elements in ERAM?	Not At All	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal

Part 2. Training

For each element, how effective was the training provided?

1. DESIREE	Not At All Effective	①②③④⑤⑥⑦⑧⑨⑩	Extremely Effective
2. Data Comm	Not At All Effective	①②③④⑤⑥⑦⑧⑨⑩	Extremely Effective

3. Traffic Based Flow Management (TBFM)	Not At All Effective	①②③④⑤⑥⑦⑧⑨⑩	Extremely Effective
4. Separation Management (SepMan)	Not At All Effective	①②③④⑤⑥⑦⑧⑨⑩	Extremely Effective

Part 3. Symbol Complexity

1. Were you able to determine the transfer of communication status of each aircraft?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
2. Could you determine the difference between equipped and non- equipped Data Comm aircraft.	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
3. Were you able to determine track ownership of the datablocks?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	Always
4. Could you tell which aircraft were ADS-B equipped?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	Always
5. Were the safety alerts easy to see in the data blocks?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	Always
6. Could you determine the meaning of different colored symbols?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	Always

Part 4. Traffic management

1. Was communications delay an issue during the simulation?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	Extremely
2. How difficult did you find the airborne reroute process?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	Extremely
3. How helpful did you find the Delay Countdown Timer to controlling traffic?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal

Data Comm

1. Did you find the Data Comm symbols confusing?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal

2. Were you able to determine when an aircraft was data comm equipped?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
3. Did color help with your interpretation of Data Comm?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
4. Did the increased datablock symbols cause clutter on your scope?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
5. Could you determine the status of a message you sent to an aircraft?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
6. How quickly were you able to send messages?	Not Quickly	①②③④⑤⑥⑦⑧⑨⑩	Very Quickly
7. Did you feel forced to use one input method over another (i.e keyboard vs trackball)	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
8. Did Data Comm flow with the rest of the system?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
9. Did you feel like you had to remember more in your head while using Data Comm?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal

Time Based Flow Management (TBFM)

1. Did you find the TBFM symbols confusing?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
2. Were you able see indication that your traffic goal was achieved?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
3. Did you find the extra data blocks distracting?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
4. Did the colors used in the Range Data block confuse you?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
5. How much did the Range Data Block and Delay Time Counter help you absorb delay time?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal

6. Did you find it difficult to use one input method over another (Keyboard, GUI)	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
7. Did you feel like you had to remember more in your head while under TBFM?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal

Separation Management (SEPMAN)

1. Did the system provide indication that you had acknowledged an alert?		YES/ NO	
2. Were you able to understand the symbols?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
3. Could you determine the different types of conflict by color?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
4. Did you feel you could quickly determine separation management from the conflict probe?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
5. Were you able to see the indications given to you on the FDB?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
6. Did you find conflict probe Alert view to be excessively long?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
7. Were you able to use your preferred method of input to control the Conflict Probes?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
8. Did you feel you had to remember more in your head while using the conflict probe?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
9. Did you find the conflict probe cluttered your scope?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
10. Did the conflict probe work the way you expected it too?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
11. Was the Conflict Probe color easily discernable?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	Easily Discernable

Airborne Reroute (ABRR)

1. Did the system provide indication that you had acknowledged a reroute?		YES/NO	
2. Were you able to determine the difference between a reroute and airborne reroute indication?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
3. Were you able to tell if an aircraft received a reroute?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
4. Were you able to determine the status of the route based on the color of the indication?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
5. Were you able to see the indications on the Range Data Block?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
6. Did you find TFM Quick view to be excessively long?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal
7. Did the airborne reroute clutter your scope?	Not at all	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal

E Sample Training Schedule

Sample training schedule for participant group 1, indicating training module, sector (10, 27), and team configuration (R-side alone; R-side/RA team) for each participant (01, 02, 03, 04).

			Experiment Room									
			ER3					**ER2**				
				Sector 10		**Sector 27**		**Sector 10**		**Sector 27**		
			Scenario	**R**	**D**	**R**	**D**	**R**		**R**		
Day	**Time**											Run
1 Tue	8:00 - 8:45	Introduction & Informed Consent										
	8:45 - 9:45	Airspace & Procedures Overview										
	9:45 -10:00	Break										
	10:00-10:45	Familiarization 1	F	01		02		03		04		1
	10:45-11:00	Break										
	11:00-11:45	Familiarization 2	F	02		01		04		03		2
	11:45-12:45	Lunch										
	12:45-2:30	Classroom: CTCC; CP; DC										
	2:30-3:15	Familiarization 3	F	01		02		03		04		3
	3:15-3:30	Break										
	3:30-4:15	Familiarization 4	F	02		01		04		03		4
	4:15-4:30	Caucus										
2 Wed	8:00-8:30	Classroom: TBFM										
	8:30-9:15	Familiarization 5	F	01		02		03		04		5
	9:15-9:30	Break										
	9:30-10:15	Familiarization 6	F	02		01		04		03		6
	10:15-10:30	Break										
	10:30 -11:00	Classroom: ABRR										
	11:00-11:45	Familiarization 7	F	01		02		03		04		7

	11:45-12:45	Lunch										
	12:45-1:30	Familiarization 8	F	02		01		04		03		8
	1:30-1:45	Break										
	1:45-2:30	Training 1	M	01		02		03		04		9
	2:30-2:45	Break										
	2:45-3:30	Training 2	M	02		01		04		03		10
	3:30-4:30	Caucus										
Day	**Time**											Run
3 Thu	8:00-8:30	Eye Tracking – initial set up										
	8:30-9:15	Training 3 (R&D)	M	01	02	03	04					11
	9:15-9:30	Break										
	9:30-9:45	Eye tracking set up										
	9:45-10:30	Training 4 (R&D)	M	02	01	04	03					12
	10:30-10:45	Break										
	10:45-11:00	Eye tracking set up										
	11:00-11:45	Training 5 (R&D)	M	03	04	01	02					13
	11:45-12:45	Lunch										
	12:45-1:00	Eye tracking set up										
	1:00-1:45	Training 6 (R&D)	M	04	03	02	01					14
	1:45-2:00	Break										
	2:00-2:45	Training 7	M	01		02		03		04		15
	2:45-3:00	Break										
	3:00-3:45	Training 8	M	02		01		04		03		16
	3:45-4:30	Caucus										
4 Fri	8:00-8:30	Eye Tracking set up										
	8:30-9:15	Training 9 (R&D)	M	01	02	03	04					17
	9:15-9:30	Break										
	9:30-9:45	Eye tracking set up										
	9:45-10:30	Training 10 (R&D)	M	02	01	04	03					18
	10:30-10:45	Break										
	10:45-11:00	Eye tracking set up										

	11:00-11:45	Training 11 (R&D)	M	03	04	01	02					19
	11:45-12:45	Lunch										
	12:45-1:00	Eye tracking set up										
	1:00-1:45	Training 12 (R&D)	M	04	03	02	01					20
	1:45-2:00	Break										
	2:00-2:45*	Training 13 (R&D)	M	01	04	02	03					21
	2:45-3:00	Break										
	3:00-3:45*	Training 14 (R&D)	M	01	03	04	02					22
	3:45-4:30	Caucus										

* mixed ATC experience levels, time allowing

F Sample Test Schedule

Sample test schedule for the first two test days for participant group 1, indicating scenario (M, B), position (R, RA) and sector (10, 27), and team configuration (R-side alone; R-side/RA team) for each participant (01, 02, 03, 04).

			ER3				**ER2**			
			Sector				**Sector**			
			10		**27**		**10**		**27**	
	Scenario	**Configuration**	**R**	**RA**	**R**	**RA**	**R**		**R**	
Day 1										
	M	R-side only	01		02		03		04	
	B	R-side only	01		02		03		04	
	B	R-side only	02		01		04		03	
	M	R-side only	02		01		04		03	
*	M	Mixed Experience Team	02	04	01	03				
Day 2										
	M	Teams	02	01	04	03				
	B	Teams	02	01	04	03				
	B	Teams	04	03	02	01				
	M	Teams	04	03	02	01				
*	M	Mixed Experience Team	04	01	02	03				

* = conducted if time allowed