**DOT/FAA/TC-12/62**

Federal Aviation Administration
William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

# Separation Management (SepMan 2): Human Factors Evaluations of Conflict Probe Location and Format and Workstation Display Alternatives

Carolina M. Zingale, PhD, FAA Human Factors Branch
Ben Willems, FAA Human Factors Branch
Kenneth Schulz, PhD, Hi-Tec Systems, Inc.
J. Stephen Higgins, PhD, Hi-Tec Systems, Inc.

October 2012

Technical Report



U.S. Department of Transportation
**Federal Aviation Administration**

# NOTICE

**Technical Report Documentation Page**

| 1. Report No. | 2. Government Accession No. | 3. Recipient's Catalog No. |
|---|---|---|
| DOT/FAA/TC-12/62 | | |

| 4. Title and Subtitle | 5. Report Date |
|---|---|
| Separation Management (SepMan 2): Human Factors Evaluations of Conflict Probe Location and Format and Workstation Display Alternatives | October 2012 |
| | 6. Performing Organization Code<br>ANG-E25 |
| 7. Author(s)<br>Carolina M. Zingale, PhD, FAA Human Factors Branch<br>Ben Willems, FAA Human Factors Branch<br>Kenneth Schulz, PhD, Hi-Tec Systems, Inc.<br>J. Stephen Higgins, PhD, Hi-Tec Systems, Inc. | 8. Performing Organization Report No.<br>DOT/FAA/TC-12/62 |
| 9. Performing Organization Name and Address<br>Federal Aviation Administration<br>Human Factors Branch<br>William J. Hughes Technical Center<br>Atlantic City International Airport, NJ 08405 | 10. Work Unit No. (TRAIS) |
| | 11. Contract or Grant No. |
| 12. Sponsoring Agency Name and Address<br>Federal Aviation Administration<br>Human Factors Research and Engineering Group<br>800 Independence Avenue, S.W.<br>Washington, DC 20591 | 13. Type of Report and Period Covered<br>Technical Report |
| | 14. Sponsoring Agency Code<br>ANG-C1 |

15. Supplementary Notes

16. Abstract

**Objective**: The En Route Automation Modernization (ERAM) system will continue to integrate new capabilities and features to assist controllers in managing increasingly high levels of traffic in the National Airspace System (NAS). The Federal Aviation Administration Human Factors Branch conducted a study to examine potential modifications to the system. The study included an evaluation of (a) the location and format of conflict probe notifications, (b) the use of 30-inch displays, and (c) the use of different pointing devices at en route controller workstations. This report summarizes the evaluations of the conflict probe and display alternatives. **Background**: This is the second Separation Management simulation. The first simulation investigated the automation requirements needed to assist controllers in working en route sectors capable of accommodating 3-nmi (5.56 km) and 5-nmi (9.26 km) separation standards. **Method**: Eighteen (12 retired, 6 current) en route air traffic controllers participated in the study. All participants managed high-traffic level scenarios under test conditions that varied the location and format of the conflict probe notifications on controller displays and that compared controller management of traffic using current displays and 30-inch displays. We collected system and participant performance measures and ratings of workload and performance. **Results**: We did not find significant differences across test conditions for measures of performance, efficiency, and safety. We did find that radar (R)-side participants viewed a higher proportion of notifications when only the most imminent notifications were presented on their displays. When notifications were available on the R-side display, the participants tended to select them more often from the data block than from the Conflict Alert List. In the Display Evaluation, the 30-inch displays were rated more favorably by the data (D)-side participants than by the R-side participants. **Conclusion**: Providing conflict probe notifications for imminent notifications on the R-side display and providing a 30-inch display on the D-side appear useful.

| 17. Key Words | 18. Distribution Statement |
|---|---|
| Conflict Probe<br>En Route Automation Modernization (ERAM) System<br>Java En Route Development Initiative (JEDI)<br>User Request Evaluation Tool (URET) | This document is available to the public through the National Technical Information Service, Alexandria, Virginia, 22312. A copy is retained for reference at the William J. Hughes Technical Center Library. |

| 19. Security Classification (of this report) | 20. Security Classification (of this page) | 21. No. of Pages | 22. Price |
|---|---|---|---|
| Unclassified | Unclassified | 81 | |

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

THIS PAGE IS BLANK INTENTIONALLY.

# Table of Content

# List of Illustrations

**Figures** **Page**



**Tables** **Page**

## Acknowledgments

Many people contributed to the development and execution of this simulation. We especially want to thank Gary Mueller of the Federal Aviation Administration (FAA) Human Factors Branch, Joel Grochowski of the Concepts and Systems Integration Branch, and members of the Engility Corporation programming team: Matt Bruckler, Joanne Flood, Adam Granich, Vince LoCasale, Laura Lott, Nicole Smith, Yev Tabekman, Valentina Velez, and Matt Zeits. We thank these individuals for their extensive efforts to develop and integrate the highly complex prototype display elements and functionality required for this study into the en route simulation system that allowed us to evaluate the proposed concepts. They worked closely with us to modify the design elements during simulation preparation to configure them as needed and were available to remediate any system problems that arose during shakedown and data collection. We thank Robert Bortu and Ed Little (Engility Corporation) for ensuring that network systems were configured and operating as needed to gather and store the simulation data.

We thank John DiRico ( TASC, Inc.) and Kelly Stephenson (formerly of Hi-Tec Systems), retired en route air traffic controllers, for their efforts in shaking down the system and scenarios, providing feedback on system performance and functionality, and providing training to the participants throughout the study. We thank Jim McGhee and Rich Ridgway (TASC, Inc.), retired air traffic controllers, for developing the scenarios used in this simulation.

We thank Otto Smith, John Dilks, and Wallace Daczkowski (Engility Corporation), who configured the hardware components, audio and video systems, and system recording capabilities. They ensured that these systems were operating as needed during the simulation and quickly remediated any problem that arose.

We thank Dan Johnson, computer scientist with the FAA Human Factors Branch, for his extensive efforts in organizing, filtering, and configuring the complex data output files so that we were able to analyze the data effectively. We thank Albert Macias of the Concepts and Systems Integration Branch for ensuring that the eye-tracking system was configured and operating correctly and for reducing the eye-movement data collected so that we could analyze it as needed.

We thank Kevin Hallman (TASC, Inc.) and Bob Muldoon (formerly of TASC, Inc.), who supervised the administration and data collection of the eye-tracking system during the simulation. Kevin also provided key support in ensuring that all questionnaire data were entered in a timely and efficient manner. Bob also supported the Principle Investigators in the set-up and monitoring of the simulation equipment. We thank Kate McDevitt, formerly of Hi-Tec Systems, who supported the Principle Investigators in operating the systems during the simulation and who reviewed and summarized materials during data analysis.

We thank Al Rehmann (FAA Simulation Branch) and Hal Olson (Hi-Tec Systems, Inc.) for scheduling the simulation pilots who supported this project during shakedown and testing. We thank the FAA simulation pilots: Lee Dixon, Audrey Wilmer, and Bill Hickman. We also thank the following Hi-Tec simulation pilots for supporting this study: Tom Vispisiano, Larry Pecan, Tom Weston, Andre Lonchonbom, Frank Johnson, Steve Allen, John Anastas, Steve Nogar, Paul Ratte, Henry Mosely, Antonio Georgoulianos, Jack Molynaux, Frank Greiner, and George Biggie.

THIS PAGE IS BLANK INTENTIONALLY.

# Executive Summary

The Federal Aviation Administration (FAA) Separation Management Program is sponsoring human-in-the-loop (HITL) simulations and other activities to guide decisions on future modification to the En Route Automation Modernization (ERAM) system. This is the second HITL conducted by the FAA Human Factors Branch (ANG-E25) in support of this effort. The first HITL investigated the automation requirements to support controllers in managing traffic in en route sectors capable of accommodating 3-nmi (5.56 km) and 5-nmi (9.26 km) variable separation standards (Sollenberger, Willems, DiRico, Hale, & Deshmukh, 2010). This second HITL investigated three issues with 18 (12 retired and 6 current) en route air traffic controllers: (a) location and format of conflict probe notifications, (b) alternative controller workstation displays, and (c) alternative workstation pointing devices. This report summarizes the evaluations of the conflict probe and display alternatives. The pointing device evaluation is summarized in a separate report (Higgins, Willems, Johnson, & Zingale, 2012).

The participants spent eight days at the FAA William J. Hughes Technical Center, Research Development and Human Factors Laboratory (RDHFL) in Atlantic City, NJ. Participants (four at a time) arrived at the RDHFL and worked as radar (R)-side and data (D)-side teams in some conditions and independently as R-side controllers in other conditions in high-traffic level scenarios.

For the Conflict Probe Evaluation, the participants completed 45-minute scenarios that built from moderate- to high-traffic levels in five test conditions. The first three test conditions evaluated the location of the probe notifications. The first condition simulated a baseline ERAM configuration that included the existing User Request Evaluation Tool (URET) conflict probe algorithm and presented probe notifications only on the D-side display. Two other conditions presented the conflict probe notifications on the R-side display as well as on the D-side display. One condition presented the probe notifications in the Conflict Alert List on the R-side display, and the other condition presented the notifications in both the Conflict Alert List and in the aircraft data blocks.

The two other test conditions used a modified conflict probe algorithm that presented notifications differently, depending on whether the aircraft trajectories were predicted to come within 6 nmi (11.11 km) of one another within the next 6-minute period or not. In both of these conditions, notifications were presented on both the R-side and D-side displays. However, in one condition, only the most imminent alerts (those predicted within 6 nmi/6 min) were presented on the R-side. In addition, the formats used to present the information differed between the two conditions. One condition used color-coding to indicate whether the conflict was predicted to occur within the next 6 min (red) or not (yellow), and the other condition used flashing to indicate whether the conflict was predicted to occur within the next 6 min (flashing) or not (steady state).

For the Display Evaluation, the participants worked as R-side and D-side teams in 35-minute scenarios in four test conditions. One condition reflected the current equipment configuration. The R-side participants used 2,048 x 2,048 (2K) displays and the D-sides used 1,024 x 1,024 (1K) displays. In a second condition, the R-side participants used 2K displays and the D-side participants used 30-inch displays (2,560 x 1,600 with 16 bit depth). In the third and fourth conditions, both the R-side and D-side participants used 30-inch displays. However, in one of these conditions, the D-side participants had additional features and capabilities available (e.g., macros) that allowed them to perform a wider range of actions. In the other condition, the additional functions were not available.

We evaluated measures of system and controller capacity, performance, and safety as well as subjective measures of workload and situation awareness for each scenario in each component of the simulation. We measured the number of aircraft handled, the time and distance of aircraft in the sector, the number of clearances issued, the number of ground-air communications, and the number of losses of separation in each condition. We also evaluated eye movements of the R-side participants to determine whether scan patterns differed across the different test conditions. In the Conflict Probe Evaluation, we examined the number of notifications presented, selected, and viewed by the participants. The participants also completed questionnaires at the end of each scenario and at the conclusion of the simulation to provide feedback on the different test conditions. They participated in a final debrief to further explain their responses.

We did not find any significant differences across test conditions in either the Conflict Probe Evaluation or the Display Evaluation components of the simulation for the primary capacity, efficiency, and safety measures. Eye-tracking data also did not reveal differences across the test conditions. With respect to the conflict probe notifications, we did find that the R-side participants selected and viewed more notifications when we modified the conflict probe algorithm to present only the most imminent notifications on their display. The participants' subjective impressions also indicated that they found it useful to provide conflict probe notifications on the R-side display, but only when the most imminent notifications were presented. The participants preferred having the notifications presented in the data block rather than in a list and preferred color-coding of the notifications rather than flashing. Based on these results, providing conflict probe information for more imminent alerts in the data blocks of aircraft on the R-side display appears useful.

For the Display Evaluation, the R-side participants found that the 2K display was sized and shaped more appropriately than the 30-inch display. The D-side participants found that having the additional space provided by the 30-inch displays useful, but also reported that when they also had more capabilities available on the D-side display, it was distracting and took away from their ability to provide support to the R-side. Overall, providing a 30-inch display on the D-side appears useful. However, more work is needed to determine the level of functionality that should be available on the D-side and how R-side and D-side roles and responsibilities should be allocated.

## 1. INTRODUCTION

The Federal Aviation Administration (FAA) Separation Management Program is sponsoring activities to guide decisions about future modifications to the En Route Automation Modernization (ERAM) system. The FAA Human Factors Branch (ANG-E25) is involved in these activities and has completed two human-in-the-loop (HITL) simulations to investigate new display elements and system functionality to support controller management of increasingly higher levels of air traffic in the National Airspace System (NAS). The first Separation Management (SepMan) HITL investigated automation requirements to support controller management of traffic in en route sectors capable of accommodating 3-nmi (5.56 km) and 5-nmi (9.26 km) variable separation standards (Sollenberger, Willems, DiRico, Hale, & Deshmukh, 2010).

The second HITL investigated three issues with 18 (12 retired and 6 current) en route air traffic controllers: (a) location and format of conflict probe notification, (b) alternative controller workstation displays, and (c) alternative workstation pointing devices. This report summarizes the first two components of the simulation. The pointing-device evaluation is summarized by Higgins, Willems, Johnson, and Zingale (2012). The results of the HITL will provide input to the development of requirements for ERAM.

### 1.1 Background

The ERAM system is replacing the current Display System Replacement (DSR) in the NAS Air Route Traffic Control Centers (ARTCCs). ERAM and DSR share a number of basic features. Both systems provide a radar display, trackball, and keyboard at the Radar (R)-side controller workstations as well as the Display Control (DC) View that allows access to system features and the Computer Readout Display (CRD) that allows the controller to enter commands and receive system feedback. Both systems provide the CRD on the Data (D)-side display as well as the User Request Evaluation Tool (URET). URET displays electronic flight progress strips and provides conflict probe notifications. Both systems display 3-line Full Data Blocks (FDBs) for aircraft under a controller's responsibility that includes the aircraft call sign (line 1), altitude (line 2), and computer identification (CID) and speed (line 3). Both systems also display 2-line Limited Data Blocks (LDBs) for aircraft that are not currently under responsibility that include only the call sign and altitude.

ERAM differs from DSR in several ways. ERAM provides additional data blocks for aircraft in adjacent facilities. These include (a) the Paired LDB that displays the call sign and Mode C altitude; (b) the Enhanced LDB that displays the call sign, Mode C altitude, assigned or interim altitude, and altitude nonconformance indicators; and (c) the Alternate Data Block that displays the call sign, Mode C altitude, assigned or interim altitude, altitude nonconformance indicators, position symbol, leader line, and the vector line and the Range Data Block (if selected).

ERAM provides toolbars consisting of buttons that allow the controller access to system features and functions such as storing and accessing preference settings, adjusting the display range, and so forth. The controller can remove a button from the toolbar and place it elsewhere on the display if desired. The system provides the capability for the controller to create and store macros on buttons. Each macro can initiate a series of steps with a single action. For example, the controller can store the sequence of commands required to display a J-ring around an aircraft (i.e., to show minimum separation distance) onto a macro button, thus eliminating the need to type the command string into the CRD each time the feature is used.

### 1.2 Conflict Probe Notifications

ERAM will continue to integrate new capabilities and features to support controller management of traffic. As traffic volume increases, it is anticipated that controllers will become more reliant on automation tools, such as URET, to assist them. URET looks 20 minutes into the future of aircraft flight paths to determine whether potential conflicts may occur. URET provides conflict probe notifications via color-coding to indicate the predicted proximity of aircraft at their closest point of approach (CPA). A red notification indicates that the aircraft trajectories are predicted to violate separation standards (i.e., 5-nmi lateral; 1,000-ft vertical), whereas a yellow notification indicates that the conformance boundaries surrounding the trajectory are predicted to violate separation standards. The conformance boundaries add 2.5-nmi laterally, 1.5-nmi longitudinally, and 300-ft vertically for aircraft in level flight and expand when aircraft are turning, climbing, or descending (e.g., Cale, Paglione, Ryan, Timoteo, & Oaks, 1998).

URET conflict probe notifications are numbered to indicate how many potential conflicts are predicted for an aircraft. Currently, the notifications are presented only on the D-side display in the Aircraft List (ACL) and on the Graphical Plan Display (GPD) above the aircraft call sign in the data block representation. Figure 1 provides an example of the conflict notifications on the GPD, and Figure 2 provides an example of the notifications in the ACL.

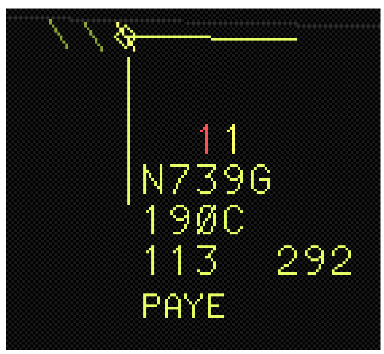


*Figure 1*. Example of URET notifications on the Graphical Plan Display.

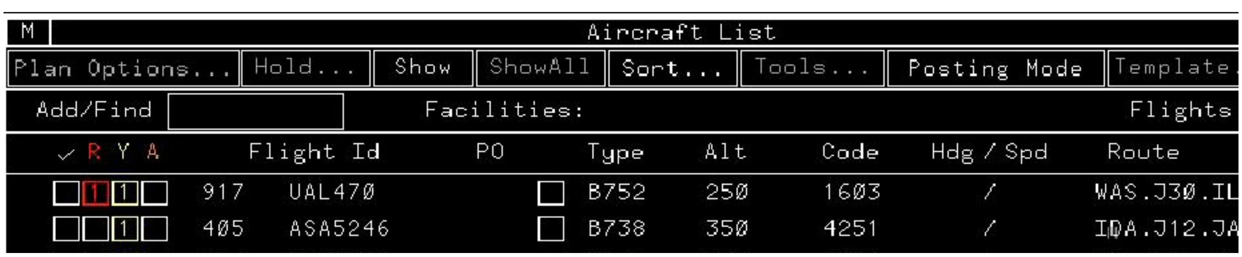


*Figure 2*. Example of URET notifications in the Aircraft List.

The extent to which controllers use URET varies widely (Cook, Duda, Willems, & Yuditsky, 2011). One reason for this may be because it is available only on the D-side display. An R-side controller working alone must divert attention from the radar display and shift his or her gaze to the D-side display to view the probe notifications. If the probe information is not readily accessible, the controller may choose not to use it. Therefore, we evaluated the effect of providing conflict probe notifications on the R-side display by presenting the notifications in the Conflict Alert List or in both the Conflict Alert List and the data block in some test conditions. However, placing additional information on the display may increase clutter to an unacceptable level. The feasibility and benefit of providing conflict notifications on the R-side display must be evaluated to determine whether they are useful and acceptable.

Another reason why URET usage may vary is that not all of the notifications have the same level of criticality, thus making some of them less valuable to the controller than others. The number of probe notifications may be reduced by modifying the way they are presented so that less imminent or less critical notifications are made less salient. To evaluate this approach, we modified the notification algorithm to indicate whether the aircraft trajectories were predicted to come within 6 nmi (11.11 km) of one another within the next 6 minutes. This strategy reduces the number of notifications presented because it no longer includes those that result when aircraft conformance boundaries are predicted to violate separation.

We used two different formats to display the modified probe notifications. One format used color-coding and the other used flashing to indicate when a conflict was predicted. The 6 minute (6 MIN) condition used color-coding, in which a red notification indicated that a conflict was predicted to occur within the next 6 minutes and a yellow notification indicated that a conflict was predicted to occur in more than 6 minutes. In this condition, the R-side displays presented only red notifications, whereas the D-side displays presented both red and yellow notifications. The red notifications were provided in Line 0 of the aircraft data block as well as in the Conflict Alert List on the radar display. Figure 3 shows an example of a conflict notification in the data block, and Figure 4 shows an example of the notification in the Conflict Alert List.


*Figure 3*. Example of conflict probe notification in the data block on the R-side display in the 6-minute condition.


*Figure 4*. Example of conflict probe notifications in the Conflict Alert List on the R-side display in the 6-minute condition.

The Flashing/Urgent condition used a flashing (on/off) symbol to indicate whether a conflict was predicted within 6 minutes, with a higher flash rate indicating higher urgency (i.e., closer time to conflict). In this condition, all conflict probe notifications were displayed via a red, rectangular indicator and a white letter A (i.e., to signify a probe alert) in Line 0 of the data block (see Figure 5). If the conflict was predicted to occur in more than 6 minutes, the indicator was presented in a steady state. If the conflict was predicted to occur within 6 minutes, the indicator flashed on and off. As the time to the predicted conflict got closer, the indicator flashed more rapidly. If the predicted conflict reached conflict alert status (i.e., predicted to occur within 2 min), the indicator and the FDB flashed at the most rapid rate and the letter "C" (for conflict) replaced the letter "A." The conflict probe notifications were also provided in the Conflict Alert List on the R-side display in this condition just as they were in the 6 MIN condition. The notifications in the Conflict Alert List did not flash.


*Figure 5*. Example of conflict probe notification in the data block in the flashing/urgent condition.

We altered the way in which the conflict probe notifications appeared for aircraft outside of the sector because sectors outside of the one used in our simulation were not staffed and coordination was not possible. Therefore, we restricted the presentation of conflict probe notifications (and the appearance of the full data block) for these aircraft until they came within 30 nmi of the sector boundary. However, if an aircraft in the sector was predicted to come into conflict with an aircraft more than 30 nmi from the sector boundary, selecting the notification for the aircraft in the sector displayed the conflict notification and full data block for the aircraft outside the sector regardless of how far it was from the sector boundary.

In both of the modified probe conditions (6 MIN and Flashing/Urgent), we provided additional functionality on the R-side and D-side displays. On the R-side, we provided conflict probe menus that displayed selectable lists of altitude, speed, and heading options and indicated by color (red, yellow, and green) whether a conflict was predicted for that selection. On the D-side, we provided a situation display instead of the GPD. The situation display functioned similarly to the R-side display and allowed the D-side participants additional interaction capabilities (e.g., routes and nodes could be selected and dragged). In these conditions, the D-side also had macros available and a Conflict Alert List. Due to the additional local functionality provided, it was necessary to establish a way for the D-side participants to interact directly with the R-side display. Therefore, when the D-side participant wanted to send a command to the R-side display, the D-side had to enter an additional command ("FWD") to designate it as such.

## 1.3 Display Alternatives

The ERAM Program Office is also making decisions about replacing displays in the en route environment with commercial-off-the-shelf (COTS) products. Choosing COTS products can provide substantial cost savings, but performance and safety criteria must be maintained, if not exceeded. Therefore, a second component of the simulation evaluated participant and system performance when alternative workstation displays were used.

We compared controller and system performance using the standard R-side, 2K (2,048 x 2,048) and D-side, 1K display with performance using 30-inch displays to determine whether differences were observed. We included four test conditions. In the first condition, we used the current R-side and D-side displays. In the second condition, the R-side used a 2K display and the D-side used a 30-inch display. The third (303R) and fourth (303S) conditions provided 30-inch displays at both the R-side and D-side positions. However, in both the 2K30 and 303R conditions, we provided the additional interaction capabilities on the d-side, as described above, for the modified conflict probe conditions. These capabilities included macros and an interactive situation display, whereas the 303S condition did not. Therefore, the 303S condition provided more limited functionality to the D-sides. We provided the two types of interaction capabilities on the D-side so that we could determine whether performance would be affected differently by the level of functionality available. We always oriented the 30-inch displays in landscape mode, and we presented the conflict probe notifications as they appeared in the 6 MIN Conflict Probe Evaluation condition.

# 2. METHOD

## 2.1 Participants

Twelve retired (11 male, 1 female) and six current[1] (5 male, 1 female) en route Certified Professional Controllers (CPCs) participated in this evaluation. The retired controllers participated in the simulation during the first 6 weeks of the simulation. The current controllers were available for the last 4 weeks of the simulation. Due to a freeze in travel funding, current controllers were not available to participate until almost the end of the 2011 fiscal year, after travel restrictions had been lifted. As a result of the travel restrictions and a requirement to complete participant travel by the end of September 2011, we were unable to recruit the full number of current controllers. We were unable to statistically compare the data obtained between the current and retired controllers in our analyses because of the low number of current controllers. Therefore, we conducted all analyses on the data for the full number of participants.

The average age of the retired controllers was 54.5 years (range: 51–59). They had an average of 28.7 years' experience (range: 22.1–34.8) controlling traffic; with, 24.1 years (range: 18.8–30.8) as CPCs for the FAA. They had an average of 24.7 years' experience (range: 20.0–30.8) in the en route environment and had last controlled traffic actively an average of 2.4 years ago (range: 0.6–4.5). Four of the retired controllers had 6.3 years' experience (range: 3–12) in the terminal environment.

[1] We planned to recruit 12 current controllers. However, due to scheduling constraints and a late cancellation, we were able to recruit only seven. The seventh controller worked with a confederate, so his data are not included in these analyses. However, he completed the pointing device evaluation and those data are summarized by Higgins, et al. (2012).

Four of the retired controllers also had an average of 3.5 years' experience (range: 1.3–5.0) as Front Line Managers (FLMs). These participants rated their current skill level as fairly high, with an average rating of 7.6 (range: 3–10) on a 10-point scale. They rated their level of motivation to participate very high, with an average rating of 9.6 (range: 8–10).

The average age of the current controllers was 43.8 years (range: 32–54). They had an average of 20.4 years' experience (range: 5.2–28.5) controlling traffic; 16.6 years (range: 2.5–25) of which were as CPCs for the FAA. They had an average of 18.4 years' experience (range: 4.5–27.3) in the en route environment and had actively controlled traffic an average of 11.2 months (range: 8–12) over the past year. Two of the current controllers also had 4 years' and 5.3 years' experience in the terminal environment, respectively. The current controllers rated their skill level as very high, with an average rating of 9.2 (range: 8–10). They also rated their level of motivation to participate very high, with all but one participant reporting it as 10.

The participants spent two weeks at the Research Development and Human Factors Laboratory (RDHFL) located at the FAA William J. Hughes Technical Center (WJHTC). The participants traveled in on Monday of the first week and traveled home on Friday of the second week. Four participants arrived at the RDHFL at a time. The participants worked in pairs as R-side and D-side controllers for conditions that required a two-person team. Nine of the participants completed the simulation as R-side controllers (6 retired, 3 current). The others participated as D-side controllers in configurations that required teams. The D-side participants completed the pointing device evaluation (Higgins et al., 2012) when the R-side participants worked independently.

### 2.2 Airspace

We used generic, high-altitude sectors in Genera center. The generic sectors in Genera center (ZGN) were originally designed by researchers and Subject Matter Experts (SMEs) at the RDHFL (Guttman & Stein, 1997). Guttman and Stein found that controllers considered the generic airspace to be realistic and their performance in the generic airspace to be comparable to their performance in real airspace. Using generic airspace allows researchers to extrapolate results across different airspaces without having to be concerned that some participants are more familiar with the airspace than others.

We combined two high-altitude (FL240-FL600) sectors (ZGN08 and ZGN22) to create a longer north-south oriented sector for use in this simulation (see Figure 6). We used the combined sectors in all components of the simulation to minimize training time. The long sector allowed us to better evaluate the feasibility of a 30-inch display monitor in landscape orientation in the display evaluation component of the simulation. Arrival routes to the primary airport, Genera, flowed in a general southbound direction, and departure routes flowed in a general northbound direction. One portion of the airspace in ZGN22 was assumed to have surveillance capabilities to allow 3 nmi (5.56 km) lateral separation of aircraft. That airspace is designated by a semicircle in Figure 6. That standard was in effect in that area in all scenarios. All other areas of the sector required 5-nmi (9.26 km) lateral separation standards.

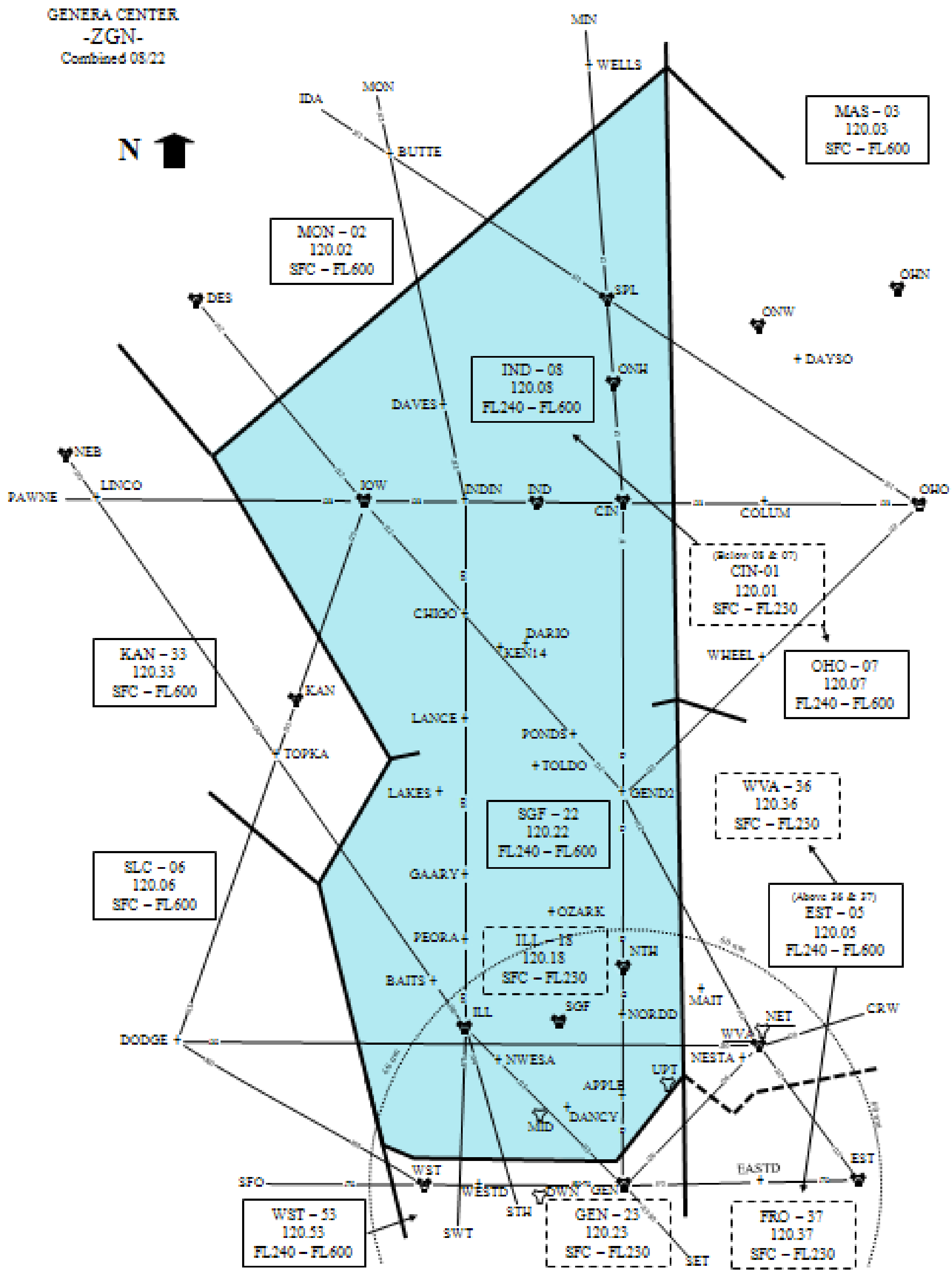


*Figure 6*. Genera center combined Sectors 08 and 22.

The Air Traffic Control (ATC) SMEs developed Standard Operating Procedures (SOPs) and Letters of Agreement (LOAs) for use in the scenarios.

- The LOAs stated
    a. that Genera (GEN) arrivals entering Sector 08 from Sector 02 shall be established on routing that proceeds over NWESA,
    b. that the GEN approach terminal area satellite arrivals entering Sector 08 from Sector 02 shall be established on routing that proceeds over SGF, and
    c. that unless otherwise coordinated, overflights on J1 entering Sector 08 from Sector 02 shall cross the 02/08 common boundary established at westbound altitudes.
- The SOPs stated
    a. that GEN arrivals entering Sector 22 from Sector 33 shall be established on routing that proceeds over NWESA and cross the 33/22 common boundary, level, at or below FL280 (Sector 22 assumes control of these aircraft on initial contact);
    b. that GEN APCH terminal area satellite arrivals entering Sector 22 from Sector 33 shall be established on routing that proceeds over SGF and cross the 33/22 common boundary, level, at or below FL280. Sector 22 assumes control of these aircraft on initial contact;
    c. that Sector 22 shall ensure GEN arrivals entering Sector 18 are cleared via routing that proceeds over NWESA direct GEN (Sector 18 assumes control of these aircraft on initial contact); and
    d. that Sector 22 shall ensure GEN APCH terminal area satellite arrivals entering Sector 18 are cleared via routing that proceeds over SGF direct to their destination (Sector 18 assumes control of these aircraft on initial contact).

## 2.3 Scenarios

We developed four high-traffic level test scenarios for use in the simulation. Traffic patterns were initially based on those obtained from Miami airspace. The scenarios began with about 10 aircraft in the sector, and then built to about 16–18 aircraft by 20 min, and had about 25–30 aircraft by 40 min. We counterbalanced the order of scenario presentation across conditions and participants. All of the test scenarios were identical except that the call signs of the aircraft differed among them. Using the same scenario in all test conditions is a common practice in simulations because it allows researchers to attribute the results observed to differences between the test conditions rather than to differences between scenarios.

We also developed four practice scenarios that had comparable traffic levels to the test scenarios but were not identical to them. The practice scenarios were identical to one another except that the call signs of the aircraft differed among them. We counterbalanced the order of the practice scenarios during training and across conditions.

We also developed one low-traffic level scenario to acclimate the participants to the airspace and procedures at the start of the simulation. This scenario had about 4 aircraft in the sector at the start of the scenario, and then built to about 12 aircraft by 15 min, and stayed at that level for the remainder of the scenario.

## 2.4 Equipment

We conducted this simulation at the RDHFL. We conducted the conflict probe and display evaluation components of the simulation in Experiment Room (ER) 3, which included two R-side and two D-side controller workstations. Video and audio equipment recorded the participants' communications and actions during the simulation so that we could review the simulation as needed when we analyzed the data. The simulation pilot workstations were located in a separate room of the RDHFL.

For the Conflict Probe Evaluation, the R-side controller workstations were equipped with a high-resolution, 2K (2,048 x 2,048) radarscope, keyboard, CRD, and trackball. The D-side workstations were equipped with a 21-inch display inset on a 30-inch monitor (EIZO ColorEdge CG303W) to reflect the current D-side 1K display. For the Display evaluation, we used a 2K or 30-inch monitor on the R-side and a 21-inch display (1K) or a 30-inch monitor on the D-side depending on the test condition. The D-side displays were mounted on adjustable arm brackets so that they could be positioned towards the R-side when R-side participants were working alone.

We used the Distributed Environment for Simulation, Rapid Engineering, and Experimentation (DESIREE) to simulate the Baseline ERAM system and then added capabilities to that system to emulate advanced capabilities and features required in the test conditions. We used the Target Generation Facility (TGF) that generated radar track and data block information based on stored flight plans. The TGF accepts entries from the simulation pilot workstations to control aircraft maneuvers that DESIREE displays on the controller workstations. We used the Java En Route Development Initiative (JEDI)/URET prototype developed by the MITRE Corporation that functions similarly to URET that is used in the field, but can be used with DESIREE. We were not able to implement airport filters into the simulation, so conflict notifications on arrival aircraft were more prevalent in the simulation than they would be in the field.

The participants completed one 45-minute, high-traffic level test scenario under each condition in the Conflict Probe Evaluation and completed one 35-minute, high-traffic level test scenario under each condition in the Display Evaluation condition. During each scenario, the participants used the Workload Assessment Keypad (WAK) to provide workload ratings (Stein, 1985). The WAK consists of a touch-panel display with 10 numbered buttons. The WAK prompts the participant to provide a subjective workload rating by using auditory and visual signals. During the prompt, the numbered buttons on the device illuminate and the device emits a brief tone. The participants indicated their current level of workload by pressing one of the numbered buttons. A rating of 1 indicated *very low workload* and a rating of 10 indicated *very high workload*. The buttons remained illuminated for the duration of the response period (20 sec) or until the participant made a response, whichever occurred first. The system recorded the rating at each interval for later analysis. If no response was made, the system recorded a missing data code.

### 2.4.1 Oculometer

We used an oculometer that consists of an eye- and head-tracking system (Applied Science Laboratories, 1991) to record Point of Gaze (POG) and pupil diameter by using near-infrared reflection outlines from the pupil and the cornea. Willems and Truitt (1999) and Willems, Allen, and Stein (1999) provide a detailed description of the hardware and the software used by this system. Willems et al. reported that exposure to infrared illumination while wearing the oculometer is less than 4% of the intensity of that experienced when outside on a sunny day.

We used the oculometer to record eye movements of the R-side participants as they worked traffic during the final practice scenario so that they could become acclimated to the device and in all of the test scenarios. We were particularly interested to determine whether participants viewed the D-side displays less frequently when they had the conflict probe information available on the R-side display in the Conflict Probe Evaluation. We also wanted to determine whether there were any measureable differences in the scanning behavior of the participants when they used the 2K and 30-inch displays that could suggest differences in how information was obtained.

### 2.4.2 Communication Systems

We used a simulated Voice Switching and Control System (VSCS) to enable voice communications. The simulated VSCS had communication links between the participants and the simulation pilots and Push-To-Talk (PTT) recording capability. The equipment monitors and records the times and durations of PTT activity. All of the communications made in the Conflict Probe Evaluation and Display Evaluation components of the simulation were made by voice.

## 2.5 Materials

### 2.5.1 Informed Consent Statement

Each participant read and signed an informed consent statement before the experiment (Appendix A). The informed consent statement described the purpose of the study and the rights and responsibilities of the participants, including that their identities and data would be kept confidential and anonymous. We used slightly different versions of this form for the retired and current controllers.

### 2.5.2 Background Questionnaire

Each participant completed a Background Questionnaire before the experiment (Appendix B). The Background Questionnaire contained questions regarding age, gender, and level of ATC experience. We used slightly different versions of this form for the retired and current controllers.

### 2.5.3 Post-Scenario Questionnaire

After completing a scenario for each test condition, the participants completed a Post-Scenario Questionnaire (PSQ; see Appendix C). One section of the questionnaire contained general items that asked the participants to provide ratings about their performance, workload, situation awareness, and other general aspects of the simulation using rating scales that ranged from 1 (*poor or extremely difficult*) to 10 (*excellent or extremely easy*).

Another section of the questionnaire contained items that pertained to the Conflict Probe Evaluation conditions. The participants used rating scales that ranged from 1 (*poor*) to 10 (*excellent*) to indicate the effectiveness of the location and format of the conflict probe notifications in each of the Conflict Probe Evaluation test conditions. Another section of the questionnaire contained items that pertained to the Display Evaluation conditions. The participants used rating scales that ranged from 1 (*completely disagree*) to 10 (*completely agree*) to indicate the extent to which they agreed with statements about the presentation of information on the display (e.g., *"the displays were uncluttered"*). The participants completed the relevant section of the questionnaire after completing each test scenario. The participants also had the opportunity to provide responses to open-ended questions and to include other comments about the scenario that they considered relevant.

### 2.5.4 Exit Questionnaire

The participants completed an Exit Questionnaire after completing the entire simulation (see Appendix D). One section included items that pertained to general issues about the simulation. The participants used rating scales that ranged from 1 (*not at all*) to 10 (*extremely or a great deal*) to indicate the extent to which they found aspects of the simulation realistic, the training effective, and the extent to which the data collection equipment (e.g., WAK, oculometer) interfered with their ATC performance.

The Exit Questionnaire contained items that asked the participants to compare the different presentations of the conflict probe notifications. The participants provided ratings on 5-point scales to indicate the extent to which the different locations (e.g., R-side, D-side) and formats (i.e., color-coding, flashing) supported their ability to detect potential conflicts. They used scales in which a rating of 1 indicated that a task was performed *much better with color-coded alerts*, 2 indicated that it was performed *somewhat better with color-coded alerts*, 3 indicated that there was *no difference between color-coded and flashing alerts*, 4 indicated that a task was performed *somewhat better with flashing alerts*, and 5 indicated that it was performed *much better with flashing alerts*. The participants used similar rating scales to compare the effectiveness of the conflict probe location between the D-side and R-side displays, and between a list and the data block.

The Exit Questionnaire also contained items that asked the participants to compare the different display monitors. The participants provided ratings on 5-point scales to indicate the extent to which the different displays supported information acquisition and ATC tasks. They used rating scales in which a rating of 1 indicated that a task was performed *much better with a 30-inch monitor*, 2 indicated that it was performed *somewhat better with a 30-inch monitor*, 3 indicated that there was *no difference between a 30-inch monitor and a standard display*, 4 indicated that a task was performed *somewhat better with a standard display*, and 5 indicated that it was performed *much better with a standard display*.

The participants had the opportunity to provide responses to open-ended questions and to include other comments about the simulation that they considered relevant.

## 2.6 Research Personnel

Two Engineering Research Psychologists were responsible for the overall administration of the Conflict Probe and Display Evaluations. They conducted the briefings, and supervised the data collection activities and simulator preparation and operation. Three research assistants supported the simulation. The research assistants prepared the experimental materials and assisted in the collection of data and in the administration and operation of the simulator, data recording, and eye tracking equipment. An en route air traffic SME who was fully familiar with the airspace, tools, concepts, and procedures used in the simulation provided an introduction to the airspace and training for the participants.

Hardware and software engineers prepared all equipment including the displays and the communications systems. The engineers were on standby to assist during the simulation as needed.

Eight simulation pilots participated during shakedown and testing. Four simulation pilots managed traffic in a sector.

## 2.7 Procedure

After arriving at the RDHFL, the participants attended an introductory briefing during which the researchers summarized the purpose of the simulation, described the test conditions and procedures, and discussed participant rights and responsibilities. After signing the Informed Consent Statement (Appendix A), the participants moved into the experiment room, ER3, to begin familiarization training on the airspace, systems, and procedures. The participants worked with the introductory, low-traffic level scenario to get them acclimated to the simulation environment and to give them experience with the airspace.

The participants then began training with the practice scenarios. The participants completed at least 15 practice scenarios when working as teams and 12 practice scenarios when working as R-sides alone. Practice scenarios were about 35–45 minutes in duration. Table 1 shows a sample schedule[2] for one group of participants.

[2] The schedule also includes the Pointing Device evaluation. When the R-side participants worked alone on the designated conflict probe scenarios, the D-side participants completed the pointing device evaluation in another experiment room at the RDHFL.

## Table 1. Sample Schedule of Events

**Week 1**

<table>
<tr><th>Time</th><th>Tuesday</th><th colspan="2">Wednesday</th><th colspan="2">Thursday</th><th>Friday</th></tr>
<tr><td>8:00</td><td>Arrival at Security Operations Center</td><td colspan="2">Daily Briefing</td><td colspan="2">Daily Briefing</td><td>Daily Briefing</td></tr>
<tr><td>8:30</td><td>Introduction,<br>(Forms, Airspace, LOA/SOP, etc.)</td><td colspan="2">Practice CP2 (1)</td><td>Practice CP4 (1)</td><td>Practice PD 2 (1)</td><td>Practice CP6 (1)</td></tr>
<tr><td>10:00</td><td rowspan="2">Lab Familiarization & Initial Practice</td><td colspan="2">Practice CP2 (2)</td><td>Practice CP4 (2)</td><td>Practice PD2 (2)</td><td>Practice CP6 (1)</td></tr>
<tr><td>11:00</td><td colspan="2">Practice CP2 (3)</td><td>Practice CP4 (3)</td><td>Test PD2</td><td>Practice CP6 (2)</td></tr>
<tr><td>12:00</td><td>Lunch</td><td colspan="2">Lunch</td><td colspan="2">Lunch</td><td>Lunch</td></tr>
<tr><td>1:00</td><td>Practice CP1 (1)</td><td>Practice CP3 (1)</td><td>Practice PD1</td><td>Practice CP5 (1)</td><td>Practice PD3 (1)</td><td>Practice CP7 (1)</td></tr>
<tr><td>2:00</td><td>Practice CP1 (2)</td><td>Practice CP3 (2)</td><td>Test PD1</td><td>Practice CP5 (2)</td><td>Practice PD3 (2)</td><td>Practice CP7 (2)</td></tr>
<tr><td>3:00</td><td>Practice CP1 (3)</td><td>Practice CP3 (3)</td><td></td><td>Practice CP5 (3)</td><td>Test PD3</td><td>Practice CP7 (3)</td></tr>
<tr><td>4:00</td><td>Caucus</td><td colspan="2">Caucus</td><td colspan="2">Caucus</td><td>Caucus</td></tr>
</table>

**Week 2**

<table>
<tr><th>Time</th><th>Monday</th><th colspan="2">Tuesday</th><th colspan="2">Wednesday</th><th>Thursday</th></tr>
<tr><td>8:00</td><td>Daily Briefing</td><td colspan="2">Display Alternative 1</td><td colspan="2">Daily Briefing</td><td>Daily Briefing</td></tr>
<tr><td>8:30</td><td>Refresh CP1 (1)</td><td colspan="2">9:00 –<br>Display Alternative 2</td><td>Refresh CP4 (1)</td><td>Practice PD4 (3)</td><td>Refresh CP6 (1)</td></tr>
<tr><td>10:00</td><td>Refresh CP1 (2)</td><td colspan="2">Display Alternative 3</td><td>Refresh CP4 (2)</td><td>Test PD4</td><td>Refresh CP6 (2)</td></tr>
<tr><td>11:00</td><td>Test CP1</td><td colspan="2">Display Alternative 4</td><td>Test CP4</td><td>Practice PD5 (1)</td><td>Test CP6</td></tr>
<tr><td>12:00</td><td>Lunch</td><td colspan="2">Lunch</td><td colspan="2">Lunch</td><td>Lunch</td></tr>
<tr><td>1:00</td><td>Refresh CP2 (1)</td><td>Refresh CP3 (1)</td><td>Practice PD4 (1)</td><td>Refresh CP5 (1)</td><td>Practice PD5 (2)</td><td>Refresh CP7 (1)</td></tr>
<tr><td>2:00</td><td>Refresh CP2 (2)</td><td>Refresh CP3 (2)</td><td>Practice PD4 (2)</td><td>Refresh CP5 (2)</td><td>Practice PD5 (3)</td><td>Refresh CP7 (2)</td></tr>
<tr><td>3:00</td><td>Test CP2</td><td>Test CP3</td><td></td><td>Test CP5</td><td>Test PD5</td><td>Test CP7</td></tr>
<tr><td>4:00</td><td>Caucus</td><td colspan="2">Caucus</td><td colspan="2">Caucus</td><td>Final Briefing</td></tr>
</table>

*Note*. SOP = Standard Operating Procedure; LOA = Letter of Agreement; CP = Conflict Probe Evaluation; PD = Pointing Device Evaluation.

## 2.8 Experimental Design and Analysis

The sections below describe the experimental design and data analysis conducted for the Conflict Probe Evaluation and the Display Evaluation components of the simulation, respectively.

### 2.8.1 Experimental Design and Analysis: Conflict Probe Evaluation

We evaluated the effect of the location of the conflict probe notifications when R-side controllers worked alone and when R-side and D-side participants worked in teams. We included three experimental conditions for the R-side participants working alone. In the Baseline condition, the conflict probe notifications were presented only on the D-side display. In the Baseline+ condition, the conflict probe notifications were also presented in the Conflict Alert List on the R-side display. In the Baseline++ condition, the conflict probe notifications were presented in the Conflict Alert List as well as in line 0 of the data block above the call sign on the R-side display. In all conditions, the notifications on the D-side were presented in the ACL and above the call sign in the data blocks of the aircraft on the GPD as they currently are in the field.

Due to scheduling constraints, we only included the Baseline and Baseline++ conditions when the participants worked as members of R-side and D-side teams. Table 2 provides a summary of the conditions used in the evaluation of the conflict probe location.

**Table 2. Conflict Probe Location Evaluation Test Conditions**

| | Conflict Probe Location | | |
|---|---|---|---|
| **Controller Configuration** | **Baseline** | **Baseline+** | **Baseline++** |
| R-side | D-side only | - D-side<br>- R-side<br>(Conflict Alert List) | - D-side<br>- R-side<br>(FDB & Conflict Alert List) |
| R-side & D-side Team | D-side only | | - D-side<br>- R-side<br>(FDB & Conflict Alert List) |

*Note.* FDB = Full Data Block.

We analyzed the data for the R-side participants working alone separately from the data for the R-side/D-side teams. We used one-way, repeated measures analyses of variance (ANOVAs) to evaluate the differences across the test conditions. We analyzed data from each scenario from 2 min to 44 min to allow time for the participants to acclimate to the scenario and to anticipate its conclusion. We summarized the data over 2-min segments through a scenario. We also conducted 3 (condition) x 3 (scenario interval: 2-16 min; 16-30 min; 30-44 min) repeated measures ANOVAs to evaluate the effect of increasing traffic level through the scenario for the R-sides working alone and 2 (condition) x 3 (interval) repeated measures ANOVAs to evaluate the data for the R-side/D-side teams. We compared the Baseline and Baseline++ conditions when the R-sides worked alone and when they worked in teams using a 2 (condition) x 2 (controller configuration) repeated measures ANOVA.

We evaluated the effect of the conflict probe format in another experimental design that included four test conditions: Baseline, Baseline++, 6 MIN, and Flashing/Urgent (see Table 3). We used the data obtained in the Baseline and Baseline++ test conditions from the location evaluation for these measures (shaded in Table 3). We analyzed these data using a repeated measures one-way ANOVA. We conducted 4 (condition) x 3 (interval) repeated measures ANOVAs to evaluate the effect of increasing traffic level through the scenario.

**Table 3. Conflict Probe Notification Format Test Conditions**

| | Baseline | Baseline++ | 6 MIN | Flashing/Urgent |
|---|---|---|---|---|
| **Probe Location** | D-side only | R-side & D-side | R-side & D-side | R-side & D-side |

### 2.8.2 Experimental Design and Analysis: Display Evaluation

We compared the data obtained in the Display Evaluation across four test conditions. We analyzed these data using a repeated measures ANOVA. We analyzed data from each scenario from 2 min to 34 min to allow time for the participants to acclimate to the scenario and to anticipate its conclusion. We summarized the data over 2-min segments through a scenario. We also conducted 4 (condition) x 2 (interval: 2-18 min; 18-34 min) repeated measures ANOVAs to evaluate the effect of increasing traffic level through the scenario. The participants always worked as R-side and D-side teams in this component of the simulation. Table 4 provides a summary of the conditions used in the display evaluation.

**Table 4. Display Evaluation Test Conditions**

| Display | 2K1K | 2K30 | 3030R | 3030S |
|---|---|---|---|---|
| **R-side display** | 2K | 2K | 30 inch | 30 inch |
| **D-side display** | 1K | 30 inch; *with* enhanced interaction capabilities | 30 inch; *with* enhanced interaction capabilities | 30 inch; *without* enhanced interaction capabilities |

### 2.9 Measures

We evaluated a comprehensive set of system and controller measures to investigate performance, efficiency and safety in each scenario in the Conflict Probe and Display Evaluations. Those measures included the number or aircraft handled, the aircraft time and distance in the sector, the number of clearances issued, and the number of losses of separation. We measured R-side participant eye movements to determine whether viewing behavior differed across the conditions. We were particularly interested in comparing how much the R-side participants viewed the D-side display under conditions in which they were working as members of R-side/D-side teams compared to when they worked as R-sides alone. We also measured participant reports of workload, situation awareness, and performance.

In the Conflict Probe Evaluation, we included an additional measure to obtain the participants' assessments of the utility of each probe notification. We did this by integrating a supplementary pop-up window into the scenarios that appeared whenever a probe notification was selected from any of the available notification locations (e.g., ACL, Conflict Alert List, data block) on either display. The participants were instructed to evaluate all of the conflict probe notifications in each scenario.

The pop-up window appeared in close proximity to the selected indicator and presented the call signs of the affected aircraft, the time (T) until the predicted conflict, and the lateral (L) and vertical (V) distance separating the aircraft at the time of the predicted conflict (see Figure 7). The pop-up window included rating scales so that the participant could indicate whether the notification was accurate (ACC?) and useful (USE?). We used a 3-point rating scale (L = low, M = moderate, H = high) for each variable. Once the participant provided a rating in both categories and selected the "Done" button, the window closed and the participant was able to resume working traffic. The displays remained visible and the scenarios continued when the pop-up windows were displayed. However, the participants were unable to interact with the display until they provided a rating for each item and closed the pop-up window.

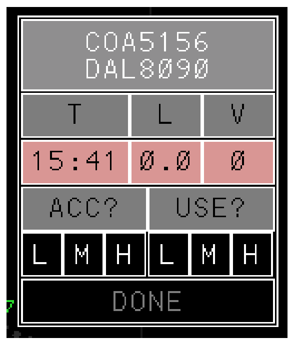


*Figure 7*. Conflict Probe Evaluation pop-up window.

## 3. RESULTS

We evaluated R-side and D-side participant data separately for variables which elicited individual measures (e.g., workload ratings) and evaluated the data as a whole when data were shared by the team (e.g., number of aircraft managed). We report the results as significant when p values were less than .05. For each analysis, when we found a violation of sphericity, we determined significance using the adjusted degree of freedom (*df*). When we found significant effects, we ran Tukey's Honestly Significant Difference (HSD) post hoc analyses to determine which pairs differed significantly from one another. Because the number of aircraft increased as the scenarios progressed, we expected the effect of interval to be significant for certain variables (e.g., number of aircraft managed). Therefore, we report the effect of scenario interval only when it interacted with test condition).

### 3.1 Number of Aircraft Managed

The sections below summarize the results obtained for the Conflict Probe Evaluation and the Display Evaluation components of the simulation, respectively.

#### 3.1.1 Number of Aircraft Managed: Conflict Probe Evaluation

We calculated the number of aircraft the participants managed across conditions. If one or more of the test conditions was more effective at supporting controller management of traffic, we would expect to find that the participants could manage more aircraft in those conditions. However, we did not find significant differences across the test conditions either when the participants worked alone (see Table 5) or when they worked as R-side/D-side teams (see Table 6) in the Conflict Probe Evaluation.

**Table 5. Mean Number (*SD)* of Aircraft Managed for R-side Participants Working Alone**

| Baseline | Baseline+ | Baseline++ |
|---|---|---|
| 57.7 (1.8) | 56.7 (2.3) | 57.8 (2.0) |

**Table 6. Mean Number *(SD)* of Aircraft Managed for R-side/D-side Teams**

| Baseline | Baseline++ | 6 MIN | Flashing/Urgent |
|---|---|---|---|
| 57.4 (1.6) | 56.8 (1.8) | 57.1 (1.8) | 57.2 (1.6) |

We also analyzed the data between the Baseline and Baseline++ conditions for participants working alone and as members of R-side/D-side teams and did not find any significant differences between these conditions.

Only the effect of scenario interval was significant. We increased traffic through our scenarios, so this was an expected result. The number of aircraft the participants managed increased as the scenarios progressed. The mean number of aircraft for which the participants were responsible at each 2-min interval is illustrated in Figure 8.

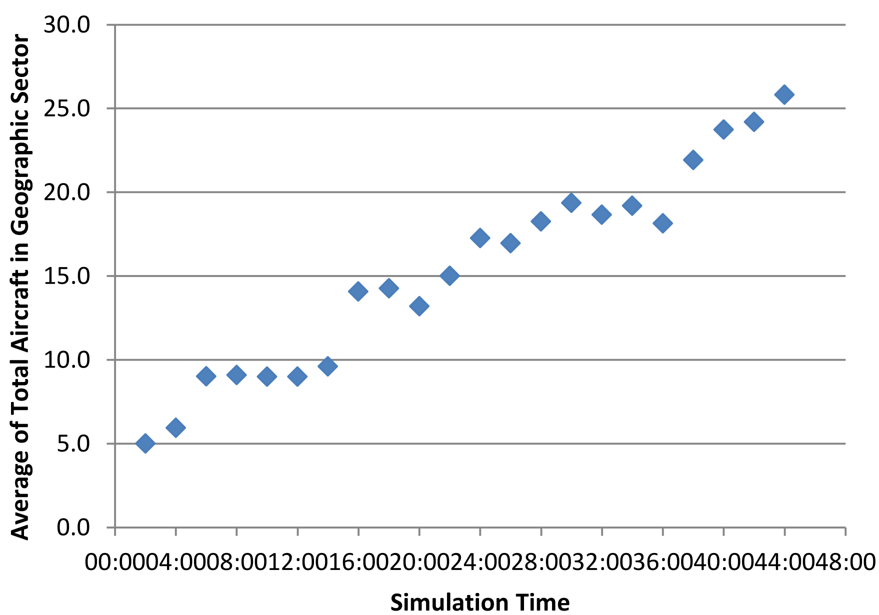


*Figure 8.* Average number of aircraft managed per 2-minute interval in the Conflict Probe Evaluation.

### 3.1.2 Number of Aircraft Managed: Display Evaluation

In the Display Evaluation, we analyzed the number of aircraft managed and did not find a statistically significant effect of condition (see Table 7). As in the Conflict Probe Evaluation, only the effect of scenario interval was significant. The overall number of aircraft managed in the Display Evaluation conditions was lower than the number managed in the Conflict Probe Evaluation due to the shorter length of the scenarios.

**Table 7. Number (*SD*) of Aircraft Managed in Each Display Evaluation Condition**

| 2K1K | 2K30 | 303S | 303R |
|---|---|---|---|
| 41.4 (1.5) | 40.5 (1.1) | 40.9 (1.4) | 41.0 (1.5) |

## 3.2 Aircraft Time and Distance in Sector

We evaluated the mean time and distance that aircraft were in the sector for each scenario in the Conflict Probe Evaluation and Display Evaluation conditions. If one or more of the test conditions better supported the participants' ability to manage the traffic, it may allow them to allocate more cognitive and system resources for developing reroute strategies and may reduce the time and distance that aircraft spend traversing the sector.

### 3.2.1 Aircraft Time and Distance in Sector: Conflict Probe Evaluation

Overall, we did not find any statistically significant differences across the Conflict Probe Evaluation conditions, whether the controllers were working alone (see Table 8) or in a team (see Table 9). The participants managed traffic similarly across the conditions.

**Table 8. Mean Time *(SD)* and Distance (*SD)* of Aircraft in the Sector for R-side Participants Working Alone**

| Time/Distance | Baseline | Baseline+ | Baseline++ |
|---|---|---|---|
| Time (sec) | 690.5 (16.1) | 695.5 (19.6) | 692.0 (13.4) |
| Distance (nmi) | 92.1 (2.7) | 93.0 (2.2) | 92.4 (2.3) |

**Table 9. Mean Time *(SD)* and Distance *(SD)* of Aircraft in the Sector for R-side/D-side Teams**

| Time/Distance | Baseline | Baseline++ | 6 MIN | Flashing/Urgent |
|---|---|---|---|---|
| Time (sec) | 688.1 (15.4) | 693.5 (17.9) | 694.0 (18.1) | 695.0 (22.1) |
| Distance (nmi) | 91.0 (1.8) | 92.0 (2.5) | 92.1 (3.1) | 92.6 (2.0) |

### 3.2.2 Aircraft Time and Distance in Sector: Display Evaluation

We evaluated the mean time and distance that aircraft were in the sector for each scenario in the Display Evaluation. We did not find any statistically significant differences across the Display Evaluation conditions. Table 10 shows the overall means and standard deviations for these data.

**Table 10. Mean Distance *(SD)* in Nautical Miles Traveled of Aircraft in the Sector for Different Display Configurations**

| Time/Distance | 2K1K | 2K30 | 303S | 303R |
|---|---|---|---|---|
| Time (sec) | 635.6 (14.5) | 640.0 (21.0) | 642.3 (16.9) | 637.5 (12.2) |
| Distance (nmi) | 70.1 (2.0) | 70.8 (2.8) | 70.7 (2.4) | 71.8 (3.0) |

## 3.3 Clearances Issued

We evaluated the number of commands issued by the participants in each scenario in the Conflict Probe and Display Evaluation conditions. If one or more of the test conditions supported the participants' ability to manage traffic more efficiently, we may find that they issued clearances differently across conditions. Such differences could indicate that the participants modified their strategies to manage the traffic differently.

### 3.3.1 Clearances Issued: Conflict Probe Evaluation

Overall, we did not find any statistically significant differences across conditions—neither when participants worked alone nor when participants worked as part of R-side/D-side teams. The average number of total commands entered in each test condition is presented in Tables 11 and 12. We did not find any statistically significant differences in the number of altitude, heading, and speed clearances issued across conditions. We did find that the effect of scenario interval was significant for the R-side participants working alone, $F(2, 16) = 32.38, p < .001$, and for the R-side/D-side participants working in teams, $F(2, 16) = 32.377, p < .001$. The number of commands increased as the scenario progressed due to the increase in traffic level.

**Table 11. Mean Number *(SD)* of Commands Issued for R-sides Working Alone**

| Baseline | Baseline+ | Baseline++ |
|---|---|---|
| 229.17 (46.8) | 225.94 (37.9) | 226.62 (39.8) |

**Table 12. Mean Number *(SD)* of Commands Issued for R-side/D-side Teams**

| Baseline | Baseline++ | 6 MIN | Flashing/Urgent |
|---|---|---|---|
| 195.72 (57.4) | 223.83 (39.2) | 195.83 (62.9) | 214.1875 (46.8) |

### 3.3.2 Clearances Issued: Display Evaluation

In this evaluation, we also examined whether participants changed the range on the displays more or less when using different sized displays. Due to the long north-south orientation of the sector and the shorter vertical size of the 30-inch monitor, we expected that the R-side participants may need to zoom in and out more frequently to obtain information on the 30-inch display to access information.

We did not find any statistically significant differences across conditions for the number of clearances issued or in the number of range adjustments made by the R-sides (see Table 13). We did find that the effect of interval was again significant, $F(1, 8) = 11.148$, $p < .05$. The number of clearances increased as the scenario progressed.

As for the D-side participants, only two of them adjusted the range (once each) during the scenarios. All of the other D-side participants adjusted the display only at the beginning of the scenario.

**Table 13. Mean Number *(SD)* of Commands Issued and Range Adjustments**

| | 2K1K | 2K30 | 303R | 303S |
|---|---|---|---|---|
| Mean Number of Commands Issued | 120.0 (39.5) | 123.0 (41.44) | 111.33 (33.42) | 122.33 (44.5) |
| Range Adjustments (R-side) | 2.6 (1.1) | 2.1 (0.3) | 2.5 (0.7) | 2.3 (0.9) |

### 3.4 Workload

The sections below summarize the results of the workload data obtained for the Conflict Probe Evaluation and the Display Evaluation components of the simulation, respectively.

#### 3.4.1 Workload: Conflict Probe Evaluation

We analyzed 22 WAK ratings within the 2-min to 44-min interval of each 45-min test scenario in the Conflict Probe Evaluation. Overall, the participants responded to 92.2% of the WAK prompts, a relatively high response rate given the frequency of the prompts. The workload ratings did not differ significantly across test conditions for the R-side participants working alone (see Table 14) or in R-side/D-side teams (see Table 15). The ratings indicated that the R-side participants found their workload fairly low to moderate on average.

**Table 14. Mean *(SD)* WAK Ratings for Participants Working as R-sides Only**

| Baseline | Baseline+ | Baseline++ |
|---|---|---|
| 3.9 (1.5) | 3.7 (1.9) | 3.7 (1.8) |

**Table 15. Mean *(SD)* WAK Ratings for R-side Participants Working as R-side/D-side Teams**

| Baseline | Baseline++ | 6 MIN | Flashing/Urgent |
|---|---|---|---|
| 3.3 (1.8) | 3.6(1.9) | 3.3 (1.8) | 3.3 (1.9) |

When we evaluated workload through the scenario, we found that the ratings increased as the scenario progressed. We calculated correlation coefficients for the WAK ratings with the number of aircraft under responsibility and found that approximately 86% of the variance of subjective R-side workload could be attributed to traffic level (see Figure 9) in each condition. The D-side participants' workload also did not differ significantly across the test conditions (see Table 16).

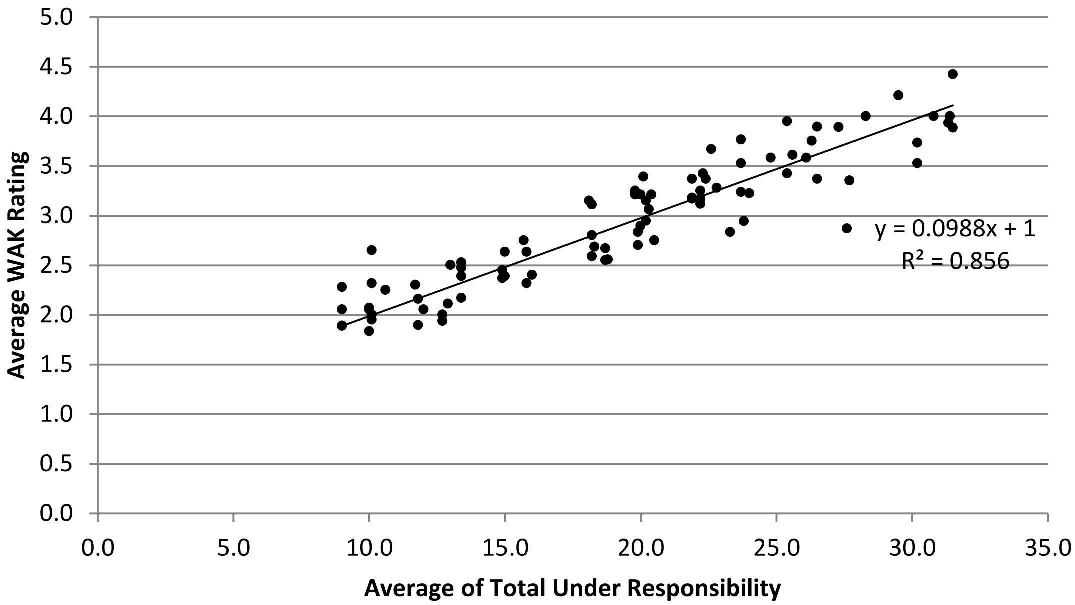


*Figure 9*. R-side participant workload ratings as a function of traffic level.

**Table 16. Mean *(SD)* WAK Ratings for D-side Participants in the Conflict Probe Conditions**

| Baseline | Baseline++ | 6 MIN | Flashing/Urgent |
|---|---|---|---|
| 3.0 (2.4) | 2.0 (1.4) | 2.5 (1.6) | 2.3 (1.5) |

However, we did find that the D-side workload ratings were affected differently than the R-side ratings across the conditions as the scenarios progressed. We correlated each R-side and D-side participant's workload ratings with traffic count in each scenario (see Table 17) and conducted a 4 (condition) x 2 (position) ANOVA on the $R^2$ values (transformed into Z-scores for the purposes of the analysis) and found a significant interaction, $F(3, 24) = 4.11$, $p = .049$. Tukey's HSD indicated that the R-side and D-side correlations between workload and traffic count did not differ significantly in the Baseline condition, but did differ in the other conditions. Factors other than traffic level influenced the D-side workload ratings in the Baseline+, the 6 MIN, and the Flashing/Urgent conditions. It is possible that when the R-side participants had probe notifications available on their displays, other aspects of the scenario influenced the D-side ratings. We examined the data to determine whether the D-side workload ratings correlated with system entries because in the 6 MIN and Flashing/Urgent conditions, the D-sides had more interaction capabilities available on their display. However, we did not find significant correlations between these measures.

**Table 17. Mean *(SD)* of Correlations ($R^2$) Between WAK Ratings and Aircraft Counts in the Conflict Probe Evaluation**

| | Baseline | Baseline++ | 6 MIN | Flashing/Urgent |
|---|---|---|---|---|
| **R-side** | 0.64 (0.4) | 0.86 (0.4) | 0.67 (0.4) | 0.79 (0.4) |
| **D-side** | 0.48 (0.4) | 0.30 (0.4) | 0.32 (0.4) | 0.27 (0.3) |

### 3.4.2 Workload: Display Evaluation

We analyzed 17 WAK ratings from within the 2-min to 34-min interval in each 35-min test scenario in the Display Evaluation. Overall, the participants responded to 94% of the WAK prompts, a relatively high response rate given the frequency of the prompts. We did not find significant differences in workload for either the R-side or D-side participants across conditions in the display evaluation (see Table 18). The average workload ratings were all less than 3, indicating fairly low levels of perceived workload.

**Table 18. Mean *(SD)* WAK Ratings for R-side and D-side Participants in the Display Evaluation**

| | 2K1K | 2K30 | 303S | 303R |
|---|---|---|---|---|
| **R-side** | 2.7 (1.4) | 2.6 (1.3) | 2.6 (1.2) | 2.8 (1.3) |
| **D-side** | 2.4 (0.4) | 2.1 (0.446) | 2.3 (0.396) | 1.9 (0.354) |

We also correlated the R-side and D-side participants' workload ratings with traffic count in each scenario (see Table 19) as we did for the Conflict Probe Evaluation. We conducted a 4 (condition) x 2 (position) ANOVA on the $R^2$ values (transformed into Z-scores for the purposes of the analysis) and found that the R-side and D-side correlations differed significantly from one another, $F(1, 8) = 6.5$, $p = .034$. The R-side participant workload ratings correlated more with traffic count than did the D-side ratings. The effect of condition was also significant, $F(3, 24) = 3.4$, $p = .03$, indicating higher correlations of workload with traffic level in the 2K1K configuration. However, Tukey's HSD showed this difference to be only marginally significant and only between the 2K1K and 303R conditions.

**Table 19. Mean *(SD)* of Correlations ($R^2$) Between WAK Ratings and Aircraft Counts in the Display Evaluation**

| | 2K1K | 2K30 | 303S | 303R |
|---|---|---|---|---|
| **R-side** | .64 (.52) | .48 (.34) | .50 (.40) | .63 (.52) |
| **D-side** | .38 (.34) | .20 (.26) | .14 (.20) | .26 (.29) |

### 3.5 Mean Handoff Acceptance Time

We evaluated the time at which participants accepted handoffs of aircraft relative to the time the aircraft crossed the sector boundary. We examined this measure as a potential indicator of cognitive workload and/or situational awareness. The participants may accept handoffs more promptly if workload is less demanding, whereas they may be unable to accept handoffs in a timely fashion when they are too busy with other aspects of managing traffic. If one or more of the test conditions provided information to allow the participants to work more efficiently, we would expect to find a significant difference in this measure across conditions.

### 3.5.1 Mean Handoff Acceptance Time: Conflict Probe Evaluation

We did not find any statistically significant differences in mean handoff acceptance time across conditions in the Conflict Probe Evaluation whether the participants were working alone (Table 20) or as a team (Table 21). The participants were equally efficient in accepting handoffs across each of the conditions. We did find that the effect of scenario interval was significant. Handoffs were not accepted as quickly as the scenario progressed for the R-side participants working alone, $F(2, 16) = 24.5$, $p < .001$, and for the R-side/D-side teams, $F(2, 14) = 17.4$, $p < .001$. This provided an indication that as traffic levels increased, it became more difficult for the participants to complete tasks in a timely manner.

**Table 20. Mean *(SD)* Time (sec) From Handoff Acceptance to Aircraft Entry Into the Sector When Working as an R-side Only**

| Baseline | Baseline+ | Baseline++ |
|---|---|---|
| 175.4 (11.7) | 174.7 (17.7) | 175.2 (7.7) |

**Table 21. Mean *(SD)* Time (sec) From Handoff Acceptance to Aircraft Entry Into the Sector When Working as an R- and D-side Teams**

| Baseline | Baseline++ | 6 MIN | Flashing/Urgent |
|---|---|---|---|
| 178.4 (12.1) | 174.0 (17.4) | 174.8 (10.5) | 174.7 (4.8) |

### 3.5.2 Mean Handoff Acceptance Time: Display Evaluation

Overall, we did not find any statistically significant differences in mean handoff acceptance time across conditions in the Display Evaluation. We had considered that the participants may accept handoffs at different times when using different size displays because they configured the displays differently to accommodate the long north-south sector. For example, if the participants used a wider range to accommodate the sector when using the 30-inch monitors, they may not have taken handoffs of aircraft from the north (most of the arrival traffic) until they got closer to the north sector boundary. A summary of the handoff-time offset data is presented in Table 22.

**Table 22. Mean *(SD)* Time (sec) From Handoff Acceptance to Aircraft Entry Into the Sector**

| 2K1K | 2K30 | 303S | 303R |
|---|---|---|---|
| 176.6 (15.0) | 173.0 (11.2) | 171.1 (16.5) | 174.8 (10.0) |

## 3.6 Losses of Separation

We evaluated losses of separation in each of the test scenarios. Losses of separation occurred when aircraft were separated by less than 5 nmi (9.26 km) or 3 nmi (5.56 km) laterally, depending on their location in the sector, and 1,000 ft (304.8 m) vertically. We eliminated any losses of separation that occurred outside of the sector or not under participant responsibility. We eliminated any losses of separation that were shorter than a single sweep of the radar (12 sec) because the participants would not have been able to detect changes in aircraft position between radar updates. We also did not consider aircraft to have lost separation if they were separated by 900 ft to 1,000 ft (274.3 m to 304.8 m) vertically because information is not available to the controller to indicate separations of less than 100 ft (30.5 m). The ATC SME and one of the researchers evaluated any remaining separation violations to determine whether the occurrences resulted because of a system error or simulation pilot error not attributable to the participant.

We found only one loss of separation across all of the test conditions in both the Conflict Probe and Display Evaluations. This event occurred in the Baseline condition (R-side/D-side team) of the Conflict Probe Evaluation in the 5-nmi (9.26 km) separation area in which one aircraft was overtaking another in level flight. This loss of separation occurred during the middle of the scenario (27 min) when the traffic level was fairly high and lasted for 36 sec. Because the loss of separation occurred in the Baseline condition, there was no reason to suspect that it was caused by anything new included in the simulation.

### 3.7 Conflict Probe Notifications

We examined the notifications in the Conflict Probe Evaluation in several ways. We calculated the number of JEDI alerts, the number of conflict probe notifications displayed for aircraft pairs predicted to be in conflict, the number of probe notifications selected by the participants, and the usefulness and accuracy ratings assigned to the selected and viewed notifications in each condition.

We encountered several system processing problems that made some aspects of the conflict probe notification data impossible to measure accurately. We conducted a series of "audits" in which we compared data in the output files to what was presented on the participant displays by reviewing the simulation video recordings. We found a number of discrepancies that led us to revise how the data output files were generated so that we could eliminate errors or omissions in the logic used to generate the output through a series of iterations. However, we still were left with discrepancies; for example, instances in which the system output data indicated that a notification was presented but did not appear on the participant's display or that indicated that a notification had been displayed and removed several times over a very short (e.g., 2–3 sec) time interval but did not correlate with what was observed on the display. These problems with determining the timing of the individual notifications made us unable to evaluate the participants' reaction times to selecting the prompts.

### 3.8 JEDI Notifications

First, we examined the data by calculating the number of probe notifications that were initiated by JEDI in each scenario in total, before the notification algorithm modified how the notifications were presented in the advanced (6 MIN and Flashing/Urgent) conditions. The average raw total number of JEDI notifications in each scenario was 106.35 ($SD$ = 8.9) and did not differ significantly across conditions for participants working alone or working in teams. We were not surprised by this result because our scenarios were identical to one another except for the aircraft call signs, and the participants did not differ significantly in the way they managed traffic across conditions.

### 3.9 Displayed and Viewed Conflict Probe Notifications

Next, we evaluated the number of conflict probe notifications displayed for unique aircraft pairs in each scenario. We expected the number of notifications to vary across conditions because of the way that we modified the algorithm in the 6 MIN and Flashing/Urgent conditions. As expected, the number of aircraft pair notifications presented on the R-side display differed significantly across conditions for the participant teams, $F(2, 16) = 500.27$, $p < .001$ (see Table 23). Tukey's HSD indicated that all of the conditions differed significantly from one another. The greatest number of notifications appeared on the R-side display in the Baseline++ condition. The modified conditions displayed fewer notifications, with the fewest displayed in the 6 MIN condition in which only conflicts predicted within 6 minutes were presented. We did not include the Baseline condition in this analysis because there were no notifications presented on the R-side display in this condition.

**Table 23. Mean (*SD*) Number of Aircraft Pair Notifications Presented on R-side Displays for R-side/D-side Teams**

| Baseline | Baseline++ | 6 MIN | Flashing/Urgent |
|---|---|---|---|
| N/A | 43.33 (3.3) | 6.22 (3.3) | 39.44 (4.3) |

Next, we evaluated the aircraft pair conflict notifications the participants viewed. The first time that a participant selected a notification prompt, the pop-up window appeared. When this occurred, the participant was unable to continue working traffic until he or she provided a rating ("low," "medium," or "high") regarding the usefulness and accuracy of that notification. We knew that a participant viewed the notification for an aircraft pair at least once by looking at this measure. We evaluated the number of viewed notifications for each participant working alone and for each team in each condition. Due to the wide variability in the number of notifications that participants viewed, we were unable to perform statistical analyses on these data in the same way we analyzed the other data in the simulation. For example, of the 247 total (summed across scenarios) notifications viewed when the R-side participants worked alone, some participants viewed many more (e.g., 84) than others (e.g., 3). For these data, we calculated the total number of notifications viewed in each condition across scenarios relative to the total number displayed per condition across scenarios, and then we calculated the overall proportion for R-side participants working alone (see Table 24) and for R-side/D-side teams (see Table 25).

**Table 24. Proportion of Total Conflict Probe Notifications Viewed in Each Condition by R-side Participants Working Alone**

| Notifications | Baseline | Baseline+ | Baseline++ |
|---|---|---|---|
| Total notifications presented | 327 | 380 | 391 |
| Total notifications viewed | 59 | 93 | 95 |
| Proportion viewed | .18 | .24 | .24 |

**Table 25. Proportion of Total Conflict Probe Notifications Viewed in Each Condition by R-side and D-side Participants Working in Teams**

| Participants | Notifications | Baseline | Baseline++ | 6 MIN | Flashing/Urgent |
|---|---|---|---|---|---|
| **R-side** | Total notifications presented | N/A | 399 | 56 | 355 |
| | Total notifications viewed | N/A | 65 | 15 | 51 |
| | Proportion viewed | N/A | .16 | .27 | .14 |
| **D-side** | Total notifications presented | 396 | 402 | 341 | 364 |
| | Total notifications viewed | 157 | 115 | 92 | 70 |
| | Proportion viewed | .40 | .28 | .27 | .19 |

The proportion of notifications that the R-side participants viewed when working alone tended to be higher in the Baseline+ and Baseline++ conditions in which the notifications were presented on the R-side display. When the participants worked in R-side/D-side teams, the R-sides viewed proportionately more notifications in the 6 MIN condition, even though there were fewer total notifications provided on their display.

We then examined the viewed notifications more closely to determine the location from which they were selected. Table 26 shows the data for the R-side participants working alone. Three of the participants did not evaluate any conflict probe notifications in the Baseline condition when they worked independently. One of those participants also did not evaluate any of the notifications in the Baseline+ condition. However, all of the participants viewed notifications in the Baseline++ condition.

**Table 26. Number of Participants Who Viewed Conflict Probe Notifications, Total Number of Notifications Viewed, and Location Selected in Each Condition for R-side Participants Working Alone**

| Notifications | Baseline | Baseline+ | Baseline++ |
|---|---|---|---|
| Number of participants who viewed notifications | 6 / 9 | 8 / 9 | 9 / 9 |
| Total notifications viewed | 59 | 93 | 95 |
| D-side display | 59 | 15 | 0 |
| R-side (Conflict Alert List) | N/A | 78 | 36 |
| R-side (Data Block) | N/A | N/A | 59 |

Eight participants evaluated notifications in the Baseline+ condition. Six of them viewed the notifications only on the R-side display via the Conflict Alert List, whereas the other two participants evaluated notifications only on the D-side. This may suggest that for those two participants, having the conflict probe notifications available only in a list on the R-side was not salient. When these two participants worked in the Baseline++ condition, their strategy changed and they selected notifications only from the R-side display via the data block indicators, suggesting that the data block notifications were more salient. In the Baseline++ condition, all of the participants selected notifications only on the R-side display.

More participants viewed notifications in the Baseline+ and Baseline++ conditions than in the Baseline condition. In the Baseline+ condition, six of the participants selected the notifications on the R-side display (84% of the notifications), whereas two participants selected the notifications only on the D-side. In the Baseline++ condition, all of the participants selected the notifications on the R-side display, and did so more often from the data block (62%) than from the Conflict Alert List.

We evaluated the usefulness and accuracy ratings that the participants made for each of the viewed notifications, although we recognize that this only represents a relatively small proportion of the notifications presented. Table 27 shows the proportion of ratings made in each category for each condition.

**Table 27. Proportion of Accuracy and Usefulness Ratings in Each Category for the Conflict Probe Notifications Made by R-side Participants Working Alone**

| Notifications | Baseline | | Baseline+ | | Baseline++ | |
|---|---|---|---|---|---|---|
| Total notifications viewed | 59 | | 93 | | 95 | |
| **Rating** | **Accuracy** | **Usefulness** | **Accuracy** | **Usefulness** | **Accuracy** | **Usefulness** |
| Low | .33 | .39 | .44 | .47 | .42 | .54 |
| Medium | .14 | .12 | .06 | .11 | .18 | .27 |
| High | .53 | .49 | .49 | .42 | .40 | .19 |

The trends in the ratings indicated that "medium" usefulness and accuracy ratings tended to be used less frequently than the "low" and "high" ratings except in the Baseline++ condition. In the Baseline++ condition, "high" usefulness ratings were given less often, possibly suggesting that when notifications were more available or salient, they were evaluated differently. Overall, however, most participants assigned the same usefulness and accuracy ratings to a notification when a pop-up window appeared. Two participants assigned different ratings to the accuracy and usefulness prompts in the Baseline condition, three participants assigned different ratings in these categories in the Baseline+ condition, and four participants assigned different ratings in these categories in the Baseline++ condition. This may also suggest that when the data were more accessible (e.g., on the R-side, in the data block), the participants gave more consideration to rating these variables.

We evaluated the data similarly for the R-side/D-side team configuration (see Table 28). As we found for the R-side participants working alone, the participant teams varied widely in the number of notifications that they viewed. One of the teams viewed only 20 notifications, whereas another team viewed 109 notifications.

**Table 28. Number of R-side/D-side Teams Who Viewed Conflict Probe Notifications, Total Number of Notifications Viewed, and Location Selected in Each Condition**

| Notifications | Baseline | Baseline++ | 6 MIN | Flashing/Urgent |
|---|---|---|---|---|
| Number of teams who viewed notifications | 7 / 9 | 9 / 9 | 9 / 9 | 9 / 9 |
| Total notifications viewed | 157 | 180 | 107 | 130 |
| D-side display | 157 | 115 | 92 | 79 |
| R-side (Conflict Alert List) | N/A | 15 | 4 | 7 |
| R-side (data block) | N/A | 50 | 11 | 44 |

Two of the teams did not view any of the notifications in the Baseline condition. The teams also differed in their approach to viewing the notifications. On one of the teams, the D-sides viewed all of the notifications and the R-sides did not view any. Overall, when the data were available on the R-side display, the notifications were more often selected via the data block than the Conflict Alert List.

We evaluated the usefulness and accuracy ratings that the R-side/D-side participants made for each of the viewed notifications. Table 29 shows the proportion of ratings made in each category for each condition.

**Table 29. Total Number of Accuracy and Usefulness Ratings of the Conflict Probe Notifications Made by R-side and D-side Participants Working in Teams**

| Notifications | Baseline | | Baseline++ | | 6 MIN | | Flashing/Urgent | |
|---|---|---|---|---|---|---|---|---|
| Total notifications viewed | 157 | | 180 | | 107 | | 130 | |
| **Rating** | **Accuracy** | **Usefulness** | **Accuracy** | **Usefulness** | **Accuracy** | **Usefulness** | **Accuracy** | **Usefulness** |
| Low | .56 | .75 | .46 | .62 | .53 | .73 | .42 | .63 |
| Medium | .18 | .06 | .21 | .17 | .22 | .08 | .29 | .13 |
| High | .26 | .19 | .33 | .22 | .24 | .19 | .28 | .24 |

The accuracy and usefulness ratings assigned to the notifications by the R-side/D-side teams were similar to those assigned by the R-side participants working alone. The "medium" usefulness and accuracy ratings tended to be used less frequently than the "low" and "high" ratings. Only about 20% to 30% of the notifications were rated as highly accurate or highly useful across conditions. If the modifications to the probe algorithm made in the 6 MIN and Flashing/Urgent conditions provided information in a more beneficial way, we would have expected to see higher ratings of accuracy and/or usefulness in those conditions compared to the Baseline conditions, but this trend was not apparent.

## 3.10 Eye Movements

### 3.10.1 Eye Movements: Conflict Probe Evaluation

We examined the number, location (R-side vs. D-side display), and duration of eye fixations. We also examined the number, duration, and distance of saccades, and the number of blinks made by the R-side participants. We analyzed the eye data from the 2-min interval to the 32-min interval in each scenario because losses of calibration among some participants required us to eliminate any data that followed. Losses of calibration typically occur later in the scenarios. They are often due to a movement of the monocle but can also occur because of other technical difficulties. We chose the 2-min to 32-min interval to ensure that the data from each of the participants was of high quality and to ensure that the measurement window would be comparable across conditions. We also explored the entire set of data (2 min to 44 min) with a smaller set of participants who had usable eye-tracking data for the duration of the scenario. Both sets of data showed a similar pattern of results.

We did not find significant differences across the test conditions for any of these measures when participants controlled traffic as R-sides alone or when they controlled with a D-side (see Tables 30 and 31). We had hypothesized that when the participants managed the traffic alone they would make more fixations on the D-side display in the Baseline condition because the probe notifications were available only on the D-side. While there was a trend in this direction, the difference was not statistically significant ($p$ = .14).

**Table 30. Conflict Probe Evaluation: Eye-Movement Data by R-side Participants Working Alone**

| | Baseline | | Baseline+ | | Baseline++ | |
|---|---|---|---|---|---|---|
| **Measurement Type** | *M* | *(SD)* | *M* | *(SD)* | *M* | *(SD)* |
| Number of fixations (total) | 3,487.6 | (647.7) | 3,551.8 | (768.4) | 3,559.4 | (395.0) |
| Duration of fixations (msec) | 464.9 | (86.7) | 400.8 | (158.5) | 371.8 | (137.1) |
| Number of fixations on R-side | 2,928.6 | (451.3) | 3,106.8 | (593.1) | 2,744.8 | (429.3) |
| Duration of fixations on R-side | 475.0 | (62.1) | 428.7 | (164.0) | 423.0 | (155.0) |
| Number of fixations on D-side | 173.8 | (212.6) | 31.0 | (60.5) | 13.6 | (16.3) |
| Number of saccades | 2,488.4 | (536.6) | 2,483.6 | (420.3) | 2,020.6 | (887.0) |
| Distance of saccades (deg. vis. angle) | 2.5 | (0.1) | 2.3 | (0.2) | 2.6 | (0.2) |
| Duration of saccades | 29.2 | (7.2) | 38.2 | (15.9) | 42.5 | (17.1) |
| Number of blinks | 600.0 | (254.9) | 615.4 | (319.2) | 626.0 | (270.0) |

**Table 31. Conflict Probe Evaluation: Eye-Movement Data by R-side Participants Working in Teams**

| | Baseline | | Baseline++ | | Flashing Urgent | | 6 MIN | |
|---|---|---|---|---|---|---|---|---|
| **Measurement Type** | *M* | *(SD)* | *M* | *(SD)* | *M* | *(SD)* | *M* | *(SD)* |
| Number of fixations (total) | 3,523.9 | (712.4) | 3,245.6 | (878.3) | 3,221.9 | (380.6) | 3,350.3 | (696.7) |
| Duration of fixations (msec) | 418.1 | (128.3) | 438.8 | (179.8) | 437.0 | (99.1) | 354.3 | (124.4) |
| Number of fixations on R-side | 2,894.6 | (673.2) | 2,834.1 | (728.6) | 2,754.8 | (416.3) | 2,415.9 | (1,022.6) |
| Duration of fixations on R-side | 356.5 | (137.0) | 445.2 | (126.4) | 469.1 | (97.5) | 463.1 | (188.0) |
| Number of fixations on D-side | 40.1 | (93.5) | 10.9 | (17.7) | 27.8 | (25.3) | 9.0 | (13.1) |
| Number of saccades | 2,213.1 | (564.0) | 2,362.5 | (548.0) | 2,292.4 | (523.1) | 2,064.6 | (880.3) |
| Distance of saccades (deg. vis. angle) | 2.5 | (0.5) | 2.4 | (0.4) | 3.0 | (1.0) | 2.5 | (0.5) |
| Duration of saccades (msec) | 36.1 | (8.8) | 43.2 | (20.9) | 61.1 | (41.9) | 62.3 | (34.5) |
| Number of blinks | 598.1 | (196.2) | 635.5 | (259.4) | 5,55.5 | (299.0) | 4,96.6 | (269.1) |

### 3.10.2 Eye Movements: Display Evaluation

We examined the number, location (R-side vs. D-side display), and duration of eye fixations; the number, duration, and distance of saccades; and the number of blinks made by the R-side participants in the 2-min to 34-min interval of the 35-min test scenarios.

We did not find significant differences across the test conditions for any of these measures (see Table 32). We had considered that because of the different sizes and shapes of the 2K (square) and the 30-inch (rectangular) display that we might find that the participants made more saccades or longer saccades when using the 30-inch display because they had a wider potential scan area. This could suggest a potentially negative effect because visual information is not acquired during a saccade; however, our data did not indicate this.

**Table 32. Display Evaluation: Eye-Movement Data**

| Measurement Type | 2K1K *M* | 2K1K *(SD)* | 2K30 *M* | 2K30 *(SD)* | 303R *M* | 303R *(SD)* | 303S *M* | 303S *(SD)* |
|---|---|---|---|---|---|---|---|---|
| Number of fixations (total) | 3,079.5 | (1,112.2) | 3,231.5 | (921.3) | 2,985.5 | (480.2) | 3,060.0 | (1,156.4) |
| Duration of fixations (msec) | 401.6 | (130.0) | 423.2 | (115.8) | 539.9 | (127.3) | 395.6 | (77.0) |
| Number of fixations on R-side | 2,446.4 | (869.7) | 2,655.4 | (910.6) | 2,416.1 | (475.3) | 2,550.1 | (999.36) |
| Duration of fixations on R-side | 453.9 | (141.9) | 467.2 | (117.7) | 597.4 | (102.1) | 434.8 | (84.9) |
| Number of fixations on D-side | 36.5 | (51.3) | 48.1 | (49.9) | 19.1 | (23.0) | 40.3 | (46.0) |
| Number of saccades | 1,906.5 | (805.1) | 2,163.5 | (880.7) | 2,153.4 | (469.5) | 1,827.5 | (320.6) |
| Distance of saccades (deg. vis. angle) | 2.3 | (0.7) | 3.1 | (1.9) | 2.5 | (0.2) | 2.7 | (0.8) |
| Duration of saccades (msec) | 29.3 | (10.3) | 66.5 | (80.1) | 40.0 | (7.6) | 59.1 | (27.6) |
| Number of blinks | 475.0 | (267.3) | 497.5 | (224.0) | 4,73.0 | (210.8) | 513.9 | (196.0) |

### 3.11 Voice Communications

We evaluated the number and duration of ground-air PTT transmissions per scenario during the 2-min to 44-min interval in the Conflict Probe Evaluation and during the 2-min to 23-min interval in the Display Evaluation. Due to system recording problems, we had missing data for two of the test scenarios in the Conflict Probe Evaluation and one of the test scenarios in the Display Evaluation. We used the mean substitution procedure to replace these data in the analysis. We eliminated any PTTs that were less than 150 msec in duration since it would not have been possible to issue a meaningful communication in that time.

#### 3.11.1 Voice Communications: Conflict Probe Evaluation

Overall, we did not find any statistically significant differences across conditions when a participant initiated communications or a pilot initiated communications, and we did not find any significant differences detected between conditions in which the participants worked alone or in a team. Only the effect of interval was significant, with the number of transmissions increasing as the scenario progressed. The result was expected given the increase in traffic level over the scenario. A summary of the PTT data is presented in Tables 33 and 34.

**Table 33. Mean Number and Duration (*SD*) of PTT Transmissions per Condition for R-side Participants Working Alone in the Conflict Probe Evaluation**

| | Baseline Mean Number | Baseline Mean Duration (msec) | Baseline+ Mean Number | Baseline+ Mean Duration (msec) | Baseline++ Mean Number | Baseline++ Mean Duration (msec) |
|---|---|---|---|---|---|---|
| **Controller** | 109.9<br>(9.3) | 3,713.2<br>(811.2) | 117.6<br>(14.3) | 3,542.3<br>(615.0) | 108.9<br>(14.7) | 3,788.3<br>(568.2) |
| **Pilots** | 128.0<br>(13.8) | 3,360.6<br>(345.5) | 137.1<br>(17.2) | 3,254.3<br>(310.7) | 134.3<br>(12.7) | 3,301.1<br>(283.7) |

**Table 34. Mean Number and Duration (*SD*) of PTT Transmissions per Condition for Participants Working in R-side/D-side Teams in the Conflict Probe Evaluation**

| | Baseline | | Baseline++ | | 6 MIN | | Flashing/Urgent | |
|---|---|---|---|---|---|---|---|---|
| | Mean Number | Mean Duration (msec) | Mean Number | Mean Duration (msec) | Mean Number | Mean Duration (msec) | Mean Number | Mean Duration (msec) |
| **Controller** | 113.7<br>(12.8) | 3,525.6<br>(592.3) | 113.0<br>(10.0) | 3,554.7<br>(627.3) | 112.4<br>(10.0) | 3,596.4<br>(662.6) | 111.8<br>(12.6) | 3,496.0<br>(627.3) |
| **Pilots** | 127.4<br>(10.6) | 3,183.0<br>(284.8) | 134.4<br>(12.5) | 3,273.4<br>(241.0) | 129.9<br>(13.7) | 3,170.5<br>(302.2) | 126.3<br>(9.5) | 3,223.6<br>(309.0) |

### 3.11.2 Voice Communications: Display Evaluation

Overall, we did not find any statistically significant differences across conditions when a participant initiated communications or a pilot initiated communications. A summary of the PTT data is presented in Table 35. Again, only the effect of interval was significant.

**Table 35. Mean Number and Duration (*SD*) of PTT Transmissions in the Display Evaluation**

| | 2K1K | | 2K30 | | 303R | | 303S | |
|---|---|---|---|---|---|---|---|---|
| | Mean Number | Mean Duration (msec) | Mean Number | Mean Duration (msec) | Mean Number | Mean Duration (msec) | Mean Number | Mean Duration (msec) |
| **Controller** | 82.4<br>(8.0) | 3,573.6<br>(598.0) | 83.0<br>(9.7) | 3,681.2<br>(864.9) | 80.6<br>(7.4) | 3,705.7<br>(645.0) | 85.7<br>(6.9) | 3,545.3<br>(636.7) |
| **Pilots** | 95.9<br>(10.4) | 3,270.7<br>(383.7) | 100.4<br>(13.6) | 3,200.4<br>(317.3) | 97.3<br>(12.5) | 3,310.6<br>(231.2) | 101.8<br>(13.4) | 3,252.8<br>(405.6) |

## 3.12 Post-Scenario Questionnaire

### 3.12.1 Post-Scenario Questionnaire: Conflict Probe Evaluation

The participants completed a PSQ at the conclusion of each of the test scenarios. The PSQ included a section on general issues about the scenario they had just completed. These included questions on the overall difficulty of the scenario, and their ATC performance, workload, and situation awareness. These items used rating scales that ranged from 1 (*extremely difficult/poor*) to 10 (*extremely easy/excellent*). The questionnaire also included open-ended items that asked the participants to comment on what aspects of the scenario were easiest to work with and what were the hardest to work with, and why. The participants were also provided the opportunity to include any additional comments or clarifications about their experience in the scenario. The comments made in response to the open-ended questions on the PSQ are included in Appendix C.

We did not find any significant differences across conditions on these general issues for either the R-side or D-side participants. Table 36 shows the means and standard deviations for these questions.

## Table 36. Mean *(SD)* for General Conflict Probe PSQ Items

| | R-sides | | | R-side / D-side Teams | | | |
|---|---|---|---|---|---|---|---|
| **Ratings for …** | **Baseline** | **Baseline+** | **Baseline++** | **Baseline** | **Baseline++** | **6 MIN** | **Flashing / Urgent** |
| 1. **Overall difficulty** of this scenario. | | | | | | | |
| R-side | 6.44 (2.5) | 6.44 (2.6) | 6.00 (2.1) | 5.33 (2.7) | 5.44 (2.4) | 5.89 (2.3) | 5.22 (2.2) |
| D-side | | | | 5.78 (2.8) | 6.56 (1.9) | 6.78 (1.8) | 5.56 (2.5) |
| 2. **Overall level of ATC performance**. | | | | | | | |
| R-side | 8.78 (1.3) | 8.56 (1.2) | 8.89 (1.1) | 8.56 (1.4) | 8.67 (0.9) | 8.67 (1.2) | 8.44 (1.6) |
| D-side | | | | 8.78 (0.8) | 9.11 (0.6) | 9.17 (0.8) | 9.11 (0.8) |
| 3. **Overall workload**. | | | | | | | |
| R-side | 6.89 (2.4) | 6.67 (2.7) | 6.78 (2.7) | 5.89 (2.6) | 6.33 (2.2) | 6.00 (2.2) | 6.89 (1.5) |
| D-side | | | | 5.11 (2.6) | 4.67 (2.5) | 5.00 (2.2) | 4.67 (2.4) |
| 4. **Workload due to communications** with pilots. | | | | | | | |
| R-side | 5.56 (2.5) | 5.78 (2.3) | 5.56 (2.7) | 5.22 (2.3) | 5.56 (2.2) | 5.67 (2.3) | 6.67 (1.9) |
| D-side | | | | | | | |
| 5. **Overall level of situation awareness**. | | | | | | | |
| R-side | 8.33 (1.3) | 8.44 (1.1) | 8.44 (1.0) | 8.56 (1.1) | 8.56 (0.9) | 8.33 (1.0) | 8.56 (1.1) |
| D-side | | | | 8.67 (0.9) | 9.00 (0.9) | 8.67 (1.6) | 8.44 (1.7) |
| 6. **Situation awareness for *current* aircraft locations**. | | | | | | | |
| R-side | 8.11 (1.5) | 8.33 (1.3) | 8.56 (1.0) | 8.11 (1.4) | 8.33 (1.0) | 8.33 (1.3) | 8.56 (1.1) |
| D-side | | | | 8.44 (1.2) | 8.67 (0.9) | 8.67 (0.9) | 8.44 (1.7) |
| 7. **Situation awareness for *projected* aircraft locations.** | | | | | | | |
| R-side | 7.89 (2.0) | 8.11 (1.6) | 8.33 (1.1) | 8.11 (1.4) | 8.33 (1.2) | 8.00 (1.7) | 8.22 (1.6) |
| D-side | | | | 8.56 (1.0) | 8.44 (0.9) | 8.56 (1.7) | 8.44 (1.8) |
| 8. **Situation awareness for potential aircraft loss-of-separation**. | | | | | | | |
| R-side | 7.89 (1.9) | 8.56 (1.0) | 8.56 (0.9) | 8.56 (1.2) | 8.56 (0.9) | 8.44 (1.2) | 8.44 (1.1) |
| D-side | | | | 9.22 (0.4) | 8.89 (0.6) | 8.56 (1.6) | 9.00 (1.0) |
| 9. **Situation awareness for potential handoff/airspace violations**. | | | | | | | |
| R-side | 8.33 (1.9) | 7.89 (2.4) | 8.33 (1.6) | 8.44 (1.6) | 8.33 (1.7) | 8.00 (1.7) | 8.33 (1.3) |
| D-side | | | | 8.44 (1.7) | 8.44 (1.4) | 7.89 (2.1) | 8.44 (1.9) |
| 10. **Performance of the simulation pilots**. | | | | | | | |
| R-side | 8.00 (1.7) | 8.22 (1.6) | 8.33 (1.2) | 8.22 (1.4) | 8.22 (1.2) | 7.56 (1.7) | 8.11 (1.2) |
| D-side | | | | | | | |

We did find significant differences on the PSQ questions that pertained directly to the conflict probe notifications (see Table 37). The participants used rating scales that ranged from 1 (*not at all*) to 10 (*extremely*) to evaluate the effectiveness of the probe location, format, and accuracy for each test scenario. When the R-side participants worked alone, their ratings about the effectiveness of the conflict probe location differed significantly across conditions, $F(2, 16) = 12.75, p = .011$. Tukey's HSD indicated that the participants rated the effectiveness of the location of the notifications in the Baseline++ condition higher than its location in either the Baseline or Baseline+ conditions. These results indicated that the participants found that providing the notifications on the radar display in both the data block and the conflict list was better than providing them only in the list on the R-side and only on the D-side. The participant questionnaire comments also supported their preference for having the notifications on the R-side display (see Appendix C for the full set of participant comments on the Conflict Probe Evaluation).

**Table 37. Mean *(SD)* for Conflict Probe PSQ Items**

| | R-sides | | | R-side / D-side Teams | | | |
|---|---|---|---|---|---|---|---|
| **Ratings for …** | **Baseline** | **Baseline+** | **Baseline++** | **Baseline** | **Baseline++** | **6 MIN** | **Flashing/ Urgent** |
| CP1. Effectiveness of the **location** (e.g., D-side, R-side, both) of the conflict probe | | | | | | | |
| **R-side | **3.89 (2.5)** | **5.89 (2.2)** | **8.67 (1.4)** | **4.11 (2.3)** | **7.44 (2.2)** | **8.11 (2.0)** | **7.56 (1.8)** |
| D-side | | | | 6.00 (2.9) | 7.00 (1.3) | 8.00 (1.2) | 7.33 1.7) |
| CP2. **Accuracy** of the probe | | | | | | | |
| R-side | 6.89 (1.6) | 6.89 (1.8) | 7.56 (1.5) | 6.89 (1.6) | 6.56 (2.0) | 7.67 (1.7) | 7.44 (1.9) |
| D-side | | | | 6.11 (2.3) | 5.89 (2.4) | 6.67 (2.7) | 7.11 1.5) |
| CP3. Effectiveness of the **format** (e.g., colors, flashing) | | | | | | | |
| **R-side | 5.44 (2.9) | 5.89 (2.6) | 7.13 (2.5) | **5.22 (2.2 )** | **7.22 (1.8)** | **8.33 (1.7)** | **7.67 (2.4)** |
| D-side | | | | 6.11 (3.2) | 5.78 (2.2) | 7.33(1.4) | 6.22 (1.6) |

*Note*. Asterisks denote statistically significant differences.

When the R-side participants worked in teams, their ratings about the effectiveness of the conflict probe location also differed significantly across conditions, $F(3, 24) = 6.76, p = .002$. Tukey's HSD indicated that the R-side participants rated the location of the notifications lower in the Baseline condition than in any of the other conditions, indicating that providing the information only on the D-side was not as effective as providing it on the R-side and D-side. The D-side participant ratings did not differ significantly across conditions.

When the R-side participants worked in teams, their ratings about the effectiveness of the conflict probe format also differed significantly across conditions, $F(3, 24) = 5.10, p = .007$. Tukey's HSD indicated that the R-side participants rated the format used in the Baseline condition lower than the formats used in the 6 MIN or Flashing/Urgent conditions. The D-side participant ratings did not differ significantly across conditions. The participant ratings regarding the accuracy of the probe did not differ significantly across conditions.

### 3.12.2 Post-Scenario Questionnaire: Display Evaluation

The participants also completed the PSQ at the conclusion of each of the Display Evaluation test scenarios (see Table 38). Only one general item, overall level of ATC performance, differed significantly across test conditions and only for the D-side participants, $F(3, 21) = 3.66$, $p = .03$ on these items. Tukey's HSD indicated that the D-side participants reported that their ATC performance was better when they used 30-inch monitors and had expanded capabilities available (2K30; 303R). The full set of participant comments on the PSQ Display Evaluation are provided in Appendix C.

**Table 38. Mean (*SD*) for General Display Evaluation Condition PSQ Items**

| | R-side / D-side Teams | | | |
|---|---|---|---|---|
| **Ratings for …** | **2K1K** | **2K30** | **3030R** | **3030S** |
| 1. Overall difficulty of this scenario. | | | | |
| R-side | 6.22 (1.8) | 6.67 (2.0) | 6.78 (1.9) | 5.78 (2.5) |
| D-side | 6.11 (2.0) | 7.44 (2.2) | 7.33 (2.2) | 6.22 (2.1) |
| 2. Overall level of ATC performance. | | | | |
| R-side | 8.44 (0.9) | 9.00 (1.3) | 8.56 (1.4) | 8.44 (1.7) |
| **D-side | 8.44 (1.5) | 8.56 (1.5) | **9.50 (0.8)** | **8.28 (1.5)** |
| 3. Overall workload. | | | | |
| R-side | 6.89 (1.8) | 7.11 (1.8) | 7.00 (2.2) | 6.67 (2.2) |
| D-side | 5.11 (2.7) | 5.22 (3.1) | 3.67 (2.3) | 4.33 (2.7) |
| 4. Workload due to communications with pilots. | | | | |
| R-side | 5.44 (2.3) | 5.89 (2.3) | 5.78 (2.4) | 5.75 (2.7) |
| D-side | | | | |
| 5. Overall level of situation awareness. | | | | |
| R-side | 8.11 (1.5) | 8.56 (1.6) | 8.11 (1.6) | 8.22 (1.6) |
| D-side | 8.67 (1.0) | 8.00 (1.5) | 8.00 (1.5) | 8.00 (1.5) |
| 6. Situation awareness for *current* aircraft locations. | | | | |
| R-side | 8.11 (2.2) | 8.33 (2.1) | 8.00 (2.1) | 7.89 (2.2) |
| D-side | 8.56 (1.3) | 7.89 (1.6) | 8.39 (2.0) | 7.56 (1.7) |
| 7. Situation awareness for *projected* aircraft locations. | | | | |
| R-side | 8.11 (2.1) | 8.11 (2.2) | 7.78 (2.2) | 8.00 (2.2) |
| D-side | 8.33 (1.5) | 7.89 (1.8) | 8.61 (1.9) | 7.89 (1.5) |
| 8. Situation awareness for potential aircraft loss-of-separation. | | | | |
| R-side | 8.78 (0.8) | 8.89 (0.9) | 8.67 (0.9) | 8.78 (1.0) |
| D-side | 8.56 (1.1) | 8.33 (1.6) | 8.61 (2.0) | 7.89 (1.5) |
| 9. Situation awareness for potential handoff/ airspace violations. | | | | |
| R-side | 8.67 (1.2) | 8.56 (1.3) | 8.22 (1.5) | 8.56 (1.3) |
| D-side | 8.00 (2.2) | 7.89 (2.1) | 8.11 (2.4) | 7.22 (2.1) |
| 10. Performance of the simulation pilots. | | | | |
| R-side | 8.78 (0.8) | 8.56 (1.1) | 8.11 (1.5) | 8.11 (1.4) |
| D-side | | | | |

As Table 39 shows, the Display Evaluation PSQ also contained statements about the readability, legibility, and so forth of the displays. Participants indicated their agreement or disagreement with the statements by using scales that ranged from 1 (*disagree*) to 10 (*agree*). The R-side participants' ratings for the appropriateness of the display size, $F(3, 18) = 8.21$, $p = .001$, and shape, $F(3, 18) = 13.32$, $p = .001$, differed significantly across conditions. Tukey's HSD indicated that the participants rated the 2K higher on both dimensions than the 30-inch display. Although almost all the mean ratings for the items in this section of the questionnaire were fairly high (nearly 7 or higher), the average ratings for the size and shape of the 30-inch monitors were lower (between 5 and 6). The R-side participants commented that the long sector did not work well with the 30-inch display oriented in landscape mode and that there was "wasted space" on the display. None of the D-side participant ratings differed significantly across conditions for any of these ratings. Their comments indicated that the 1K displays were more cluttered and the 30-inch display allowed them to have the ACL and GPD in view simultaneously.

### Table 39. Mean (*SD*) for Display Evaluation PSQ Questions

| | **R-side / D-side Teams** | | | |
|---|---|---|---|---|
| **Ratings for …** | **2K1K** | **2K30** | **3030R** | **3030S** |
| D1. The displays were uncluttered. | | | | |
| R-side | 8.29 (2.5) | 8.14 (2.1) | 7.86 (2.2) | 7.00 (2.8) |
| D-side | 6.11 (2.8) | 5.83 (3.3) | 7.25 (2.5) | 7.33 2.4) |
| D2. I could find the information I needed quickly. | | | | |
| R-side | 9.00 (1.2) | 9.00 (1.2) | 8.71 (1.1) | 8.71 (1.0) |
| D-side | 7.00 (2.2) | 7.22 (2.2) | 8.75 (0.9) | 6.78 (2.6) |
| D3. The text was easy to read. | | | | |
| R-side | 9.00 (1.0) | 9.00 (1.2) | 8.71 (1.0) | 8.00 (1.9) |
| D-side | 7.56 (2.7) | 6.00 (3.1) | 8.38 (2.3) | 7.33 2.6) |
| D4. The graphics were legible and easy to interpret. | | | | |
| R-side | 9.00 (1.0) | 8.86 (1.4) | 8.86 (1.2) | 8.43 (1.4) |
| D-side | 7.44 (1.9) | 7.00 (2.3) | 8.13 (2.3) | 6.56 2.8) |
| D5. The display was in an ideal physical position. | | | | |
| R-side | 8.57 (1.8) | 8.43 (2.5) | 7.29 (3.35) | 6.86 (3.6) |
| D-side | 7.72 (2.0) | 7.78 (1.4) | 8.63 (0.9) | 8.11 1.5) |
| D6. The display was an appropriate size. | | | | |
| **R-side | **8.57 (1.6)** | **8.43 (2.5)** | **5.86 (2.1)** | **5.86 (3.0)** |
| D-side | 6.00 (3.0) | 7.33 (2.6) | 8.50 (1.2) | 8.11 1.8) |
| D7. The display was an appropriate shape. | | | | |
| **R-side | **8.57 (1.6)** | **8.29 (2.3)** | **5.43 (2.3)** | **5.00 (2.6)** |
| D-side | 8.22 (1.2) | 7.56 (2.3) | 8.38 (1.2) | 7.72 (1.8) |

### 3.13 Exit Questionnaire and Debrief Comments

The participants completed an Exit Questionnaire at the conclusion of the simulation and provided feedback and comments during a final debriefing session. The Exit Questionnaire included a section about simulation realism, training, and the extent to which equipment interfered with ATC performance. These items used rating scales that ranged from 1 (*not at all*) to 10 (*extremely* or *a great deal*). The participants indicated that the generic airspace ($M$ = 7.44, $SD$ = 1.58), the hardware ($M$ = 8.11, $SD$ = 1.53), and the simulation software were fairly realistic ($M$ = 7.50, $SD$ = 1.20), and they rated the training provided as highly effective ($M$ = 8.94, $SD$ = 0.83). However, the traffic scenarios were rated only as moderately realistic ($M$ = 6.00, $SD$ = 1.97). The comments made during the simulation and in the debriefing session indicated that the similarity of the scenarios made events predictable.

The participants indicated that responding to the WAK ($M$ = 2.56, $SD$ = 1.69) and wearing the oculometer ($M$ = 3.00, $SD$ = 2.24) did not interfere much with their ATC performance. Although we did not include a question about the Conflict Probe Evaluation pop-up window, the participants' comments during the simulation indicated that this measure interfered somewhat with their ability to manage the traffic, and some of the participants described it as "distracting" during the debriefing session.

The Exit Questionnaire also contained items that asked the participants to compare the different conflict probe formats and locations presented during the simulation. They used 5-point scales to indicate the extent to which the different presentation formats and locations supported their ability to detect potential conflicts. A rating of 1 indicated that a task was performed *much better* with one of the options, 2 indicated that it was performed *somewhat better* with one of the options, 3 indicated that there was *no difference* between the options, 4 indicated that a task was performed *somewhat better* with the alternative option, and 5 indicated that it was performed *much better* with the alternative option. Table 40 shows the average responses to these items.

## Table 40. Mean (*SD*) Responses to Exit Questionnaire Conflict Probe Items

| Conflict Probe Format | ① | ② | ③ | ④ | ⑤ |
|---|---|---|---|---|---|
| **Differences between Current and Modified Probe** | **Much better with current probe format** | **Somewhat better with current probe format** | **No difference between current probe format and modified format** | **Somewhat better with modified format** | **Much better with modified format** |
| Detecting potential conflicts | | | | | |
| R-side | | | | 3.89 (1.5) | |
| D-side | | | | | 4.56 (.7) |
| Determining when a potential conflict was expected to occur | | | | | |
| R-side | | | | 3.78 (1.4) | |
| D-side | | | | 4.33 (.8) | |
| | | | | | |
| **Differences between Color-coded and Flashing Probe Formats [3]** | **Much better with color-coded alerts** | **Somewhat better with color-coded alerts** | **No difference between color-coded alerts and flashing alerts** | **Somewhat better with flashing alerts** | **Much better with flashing alerts** |
| Detecting potential conflicts | | | | | |
| R-side | 1.44 (.7) | | | | |
| D-side | | 2.11 (1.4) | | | |
| Determining when a potential conflict was expected to occur | | | | | |
| R-side | | 1.67 (1.0) | | | |
| D-side | | 2.11 (1.7) | | | |

(*table continues*)

[3] Two of the participants' responses to these items indicated that they preferred the flashing format to the color-coded format. Both of these participants served as D-sides. However, when we examined the comments associated with these ratings, it was clear that they misinterpreted the questions. One participant thought he was being asked to indicate whether a flashing indicator in line 0 would be preferable to a flashing data block to indicate a conflict alert (this was not displayed in the simulation). The other participant included a comment on the questionnaire about this item, saying s/he "liked the numbers above the R-side data block for the alerts," which contradicted the response selected on the rating scale. Therefore, these data were eliminated from the analysis.

**Table 40. Mean (*SD*) Responses to Exit Questionnaire Conflict Probe Items**
*(continued from previous page)*

| Conflict Probe Format | ① | ② | ③ | ④ | ⑤ |
|---|---|---|---|---|---|
| **Conflict Probe Location** | **Much better on the D-side Only** | **Somewhat better on the D-side Only** | **No difference whether on the D-side only or on both the R-side and D-side** | **Somewhat better when on both the R-side and D-side** | **Much better when on both the R-side and D-side** |
| Detecting potential conflicts | | | | | |
| R-side | | | | | 4.78 (.44) |
| D-side | | | | 3.56 (1.0) | |
| Determining when a potential conflict was expected to occur | | | | | |
| R-side | | | | 4.44 (.7) | |
| D-side | | | | 3.44 (1.1) | |
| | | | | | |
| **Conflict Probe Layout** | **Much better with the List Only on the R-side** | **Somewhat better with the List Only on the R-side** | **No Difference between the List Only on the R-side and the List + Data Block indicators** | **Somewhat better with the List + Data Block Indicators on the R-side** | **Much better with the List + Data Block Indicators on the R-side** |
| Detecting potential conflicts | | | | | |
| R-side | | | | | 4.77 (.4) |
| D-side | | | | 3.89 (.9) | |
| Determining when a potential conflict was expected to occur | | | | | |
| R-side | | | | 4.44 (.9) | |
| D-side | | | | 3.78 (1.0) | |

Exit Questionnaire ratings indicated that the participants found that having the conflict probe notifications presented on both the R-side and D-side helped them better detect and determine when a potential conflict would occur than did presentation on the D-side alone. During the debriefing session, the R-side participants commented that they rarely looked at the D-side display even when they worked alone and the probe notifications were only available on the D-side. The eye movement data we obtained supported these comments. We found that the R-side participants made only a small percentage of fixations on the D-side displays.

With respect to the probe presentation layout, the ratings indicated that displaying the information on the R-side in both the data block and the Conflict Alert List helped them detect and determine when a potential conflict was expected to occur better than when it was displayed in the Conflict Alert List alone. However, during the debriefing session, the participants commented that lists on the R-side are a "distraction." The consensus was that the conflict probe notifications were best displayed in the data block. One participant suggested that if presenting the notifications in a list were required, it should be a selectable, on/off, option.

The Exit Questionnaire ratings also indicated that the participants found that the modified probe format used in the 6 MIN and Flashing/Urgent conditions, helped them better detect and determine when a potential conflict would occur than did the current conflict probe format notifications. The modified probe alerts indicated that the aircraft trajectories were predicted to come within 6 nmi (11.11 km) of each other and whether the conflict would occur within the next 6 minutes. Of the two modified probe formats, the participants' ratings indicated that color-coding (6 MIN) helped them better detect and better determine when a conflict would occur than did flashing (Flashing/Urgent). In the debriefing session, the participants commented that the flashing indicator was "too intrusive" or "too salient." There was a clear preference to have the notifications displayed in the data block using color-coding and to show only the most imminent alerts (red) on the R-side display. The participants reported that having all alerts presented on the D-side was reasonable, however, because the D-sides would have more opportunity to evaluate less time-critical alerts.

Other comments made during the debriefing session were related to suggested improvements to the presentation of the display elements. For the conflict probe notifications, the participants wanted the display to indicate when a probe notification had already been examined. They suggested this could be done by providing a checkmark next to the entry list item or by graying out or subduing the color of the indicator in the list and in the data block. Most participants agreed that they found themselves looking at "already examined" probe notifications more often than necessary because they had no visual indication as to which notifications had been previously examined. The participants also reported that other display elements used in the simulation were too bright. Comments focused on the intensity of the route lines that were displayed when a conflict probe notification was selected and the red Urgent/Flashing indicator.

The participants commented that the conflicts predicted within their sector were generally accurate, overall, but that it was distracting to see notifications for conflicts that were predicted to occur in other sectors. This happened more frequently in our simulation than it would in the field because airport filters were not implemented in the simulation environment. This meant that the probe notifications for overflights in the sector were reported to be more useful than those for aircraft landing at Genera or one of the satellite airports. A summary of all of the written comments from the Exit Questionnaire about the Conflict Probe Evaluation is provided in Appendix D.

The Exit Questionnaire also contained a section of items that compared the display monitors. The participants provided ratings on 5-point scales to indicate the extent to which the different display monitors supported information acquisition and ATC tasks. A rating of 1 indicated that a task was performed *much better with a 30-inch monitor*, 2 indicated that it was performed *somewhat better with a 30-inch monitor*, 3 indicated that there was *no difference between a 30-inch monitor and a standard display* (either a 2K or 1K display), 4 indicated that a task was performed *somewhat better with a standard display*, and 5 indicated that it was performed *much better with a standard display*. Table 41 shows the average responses to these items.

**Table 41. Mean (*SD*) Responses to Exit Questionnaire Display Evaluation Items**

| | ① | ② | ③ | ④ | ⑤ |
|---|---|---|---|---|---|
| | **Much Better with a 30-inch monitor** | **Somewhat Better with a 30-inch monitor** | **No Difference Between 30-inch monitor and the standard display** | **Somewhat Better with a standard display** | **Much Better with a standard display** |
| Locating information on the display | | | 3.00 (1.7) | | |
| R-side | | | | 4.22 (1.1) | |
| D-side | | 1.63 (1.0) | | | |
| Scanning traffic effectively | | | 3.00 (1.7) | | |
| R-side | | | | 4.11 (.9) | |
| D-side | | 1.75 (1.4) | | | |
| Avoiding or resolving potential conflicts | | | 3.00 (1.5) | | |
| R-side | | | | 4.11 (.9) | |
| D-side | | 1.75 (1.0) | | | |
| Maintaining situation awareness | | | 3.06 (1.6) | | |
| R-side | | | | 4.11 (.9) | |
| D-side | | 1.88 (1.5) | | | |
| Managing traffic efficiently | | | 3.18 (1.5) | | |
| R-side | | | | 4.11 (.9) | |
| D-side | | 2.13 (1.4) | | | |

The mean ratings indicated that the R-side participants generally found the current 2K displays more effective in supporting their ATC tasks than the 30-inch displays, whereas the D-side participants generally found the 30-inch displays more effective. The R-side participants commented that the 30-inch displays provided too much unused space and constrained the presentation of the long, north-south sector into one area of the monitor. Their comments indicated that a portrait orientation would have been more effective for presenting this airspace configuration. The D-side participants felt more positive about having the additional space provided by the 30-inch display because it allowed them to view both the ACL and the GPD simultaneously. However, the D-side participants also commented that it was somewhat difficult to work in the advanced D-side (2K20; 303R) conditions because there were "too many toys" and that it was "distracting." One participant reported that a "different mindset" was needed to work when the advanced D-side was provided. The configuration used in these conditions also required the D-side participants to make an additional entry when they wanted to affect something on the R-side display (e.g., move a data block) that changed the tasks they are familiar with. A complete summary of all of the participants' comments written on the Exit Questionnaire is provided in Appendix D.

## 4. DISCUSSION

In this simulation, we evaluated the effect of providing conflict probe notifications on the R-side display and the effect of modifying the probe notification format. We also evaluated the utility of replacing current workstation displays on the R-side and D-side with 30-inch monitors. We measured system and controller efficiency, performance, and safety for each component of the simulation by evaluating the number of aircraft in the sector, time and distance of aircraft in the sector, number of ground-air communications, number and type of clearances, and losses of separation. Overall, we did not find any significant differences in these measures across our test conditions. We also measured eye movements of the R-side participants to determine whether scanning behavior differed across conditions and found no significant differences for this measure. We were particularly interested in whether the R-side participants made more fixations on the D-side display when they worked alone than when they worked in R-side/D-side teams and the probe notifications were available only on the D-side displays. We had hypothesized that the R-sides would look more frequently at the D-side display when working alone in the Baseline condition because the probe notifications were available only on the D-side display. Although the data trended in this direction, we did not obtain a significant result.

We evaluated workload throughout each of the scenarios, but did not find statistically significant differences across the test conditions on these ratings. We did find a strong correlation between the number of aircraft in the scenario and workload ratings for the R-side participants in all conditions, but a strong correlation between the number of aircraft and workload was found only for the D-side participants in the Baseline condition.

The R-side participants were, generally, less satisfied using the 30-inch displays than the D-side participants. Whereas the D-side participants appreciated having more space on the display to simultaneously present the ACL and GPD, the R-side participants did not find that the rectangular shape of the 30-inch display supported their ATC tasks effectively; the R-side participants preferred the standard 2K monitor. We used a long north-south sector in this simulation, and the R-sides felt that a portrait orientation would have better accommodated this sector configuration. We did not test a portrait orientation in this simulation because that configuration is not planned for field use. The D-side participants, however, did raise concerns when using the 30-inch displays in conditions that provided them more advanced capabilities, such as macros and "click and drag" functionality. The D-side participants reported that having more available local features on their display made them feel more separated from the R-side.

For the Conflict Probe Evaluation conditions, we found, as expected, that the raw number of JEDI alerts generated for each scenario did not differ significantly across conditions. We did find that when we modified the conflict probe algorithm and presented only the most imminent notifications on the R-side display, the participants viewed a higher proportion of them overall. The number of notifications viewed by the participants varied widely. However, when the notifications were presented on the R-side display, the participants tended to select and view them more often from the data block location than from the Conflict Alert List.

We found differences across the test conditions for a number of the subjective measures. Overall, the participants found the modified conflict probe to be more effective than the current probe algorithm and that the most effective format was one in which only the most imminent alerts were presented on the R-side display. The participants also found the color-coded notifications to be more effective than the flashing notifications because they indicated that flashing was too intrusive.

Overall, placement of the conflict probe notifications on the R-side appears to be acceptable to controllers as long as only the most imminent alerts are provided and the notification format is not too intrusive. However, we do not have objective data to support that placing the conflict probe notifications on the R-side display increases efficiency, capacity, or safety. Nor do we have objective data from the Display Evaluation conditions to indicate that the 30-inch display in landscape mode would be a problem for the R-side to use if working a long north-south sector.

We recommend further investigation of the preferred conflict notification using more varied scenarios with higher traffic levels. This may better encourage the use of different control strategies across test conditions that may make differences in the performance, efficiency, and safety measures apparent. We recommend that 30-inch displays be optimized for use—depending on the orientation of the sector— if possible, so that a long north-south sector such as the one used in our simulation can be optimally accommodated.

## References


Applied Science Laboratories, Inc. (1991). *4100H Eye-tracking system with head-mounted optics specifications* [Computer hardware]. Bedford, MA: Applied Science Laboratories, Inc.

Cale, M. L., Paglione, M., Ryan, H., Timoteo, D., & Oaks, R. (1998). *User Request Evaluation Tool (URET) Conflict prediction for accuracy report.* Retrieved from http://www.acy.tc.faa.gov/cpat/ccld_scen/RiskReduction/acc.../ucpa_fin/pdf

Cook, C., Duda, A., Willems, B., & Yuditsky, T. (2011). *Pre-ERAM baseline study* [Internal report]. Atlantic City International Airport, NJ: FAA William J. Hughes Technical Center. Unpublished manuscript.

Guttman, J. A., & Stein, E. S. (1997). *En route generic airspace evaluation* (DOT/FAA/CT-TN97/7). Atlantic City International Airport, NJ: Federal Aviation Administration William J. Hughes Technical Center.

Higgins, J. S., Willems, B., Johnson, D., & Zingale, C. M. (2012). *Human factors evaluation of pointing devices used by Air Traffic Controllers: Changes in physical workload and behavior*. Manuscript in preparation.

Sollenberger, R. L., Willems, B. F., DiRico, J., Hale, M., & Deshmukh, A. (2010). *Human-in-the-loop investigation of automation requirements for separation management* (DOT/FAA/TC-10/07). Atlantic City International Airport, NJ: Federal Aviation Administration William J. Hughes Technical Center.

Stein, E. S. (1985). *Air Traffic Controller workload: An examination of workload probe* (DOT/FAA/CT-TN84/24). Atlantic City International Airport, NJ: Federal Aviation Administration Technical Center.

Willems, B. F., Allen, R. C., & Stein, E. (1999). *Air Traffic Control Specialist visual scanning II: Task load, visual noise, and intrusions into controlled airspace* (DOT/FAA/CT-TN99/23). Atlantic City International Airport, NJ: Federal Aviation Administration William J. Hughes Technical Center.

Willems, B. F., & Truitt, T. R.(1999). *Implications of reduced involvement in en route Air Traffic Control* (DOT/FAA/CT-TN99/22). Atlantic City International Airport, NJ: Federal Aviation Administration, William J. Hughes Technical Center.

## Acronyms

| | |
|---|---|
| ACL | Aircraft List |
| ARTCC | Air Route Traffic Control Center |
| ATC | Air Traffic Control |
| COTS | Commercial-Off-The-Shelf |
| CPC | Certified Professional Controller |
| CRD | Computer Readout Display |
| DESIREE | Distributed Environment for Simulation, Rapid Engineering, and Experimentation |
| D-side | Data-side |
| DSR | Display System Replacement |
| ER | Experiment Room |
| ERAM | En Route Automation Modernization |
| FAA | Federal Aviation Administration |
| FDB | Full Data Block |
| GEN | Genera |
| GPD | Graphical Plan Display |
| HITL | Human-In-The-Loop |
| HSD | Tukey's Honestly Significant Difference |
| JEDI | Java En Route Development Initiative |
| LDB | Limited Data Block |
| LOA | Letter of Agreement |
| NAS | National Airspace System |
| PSQ | Post-Scenario Questionnaire |
| PTT | Push-To-Talk |
| RDHFL | Research Development and Human Factors Laboratory |
| R-side | Radar-side |
| SepMan | Separation Management |
| SME | Subject Matter Expert |
| SOP | Standard Operating Procedure |
| TGF | Target Generation Facility |
| URET | User Request Evaluation Tool |
| VSCS | Voice Switching and Control System |
| WAK | Workload Assessment Keypad |
| WJHTC | William J. Hughes Technical Center |
| ZGN | Genera Center |

# Appendix A: Informed Consent Statement

## Informed Consent Statement

Separation Management (SepMan II) Human-in-the-Loop Simulation

I, __________________________________, understand that this simulation, entitled "Separation Management (Sep Man) II: Evaluations of conflict probe location and format, display monitors, and pointing devices:" is sponsored by the Federal Aviation Administration (FAA) and is being directed by Dr. Carolina Zingale.

**Nature and Purpose:**

I have been recruited to volunteer as a participant in this simulation that will consist of three components. These components will investigate: a) the location and format of the conflict probe, b) alternatives to the existing display monitors at the radar (R-side) and data (D-side) positions, and c) three alternative pointing devices to the workstation trackball. This simulation will evaluate these issues in high traffic scenarios using a simulated En Route Automation Modernization (ERAM) system. I understand that the participants will be randomly assigned to work as either R-side or D-side controllers in some conditions. Depending on the condition, I will also be asked to wear a head-mounted oculometer to record eye movements or electromyographic (EMG) sensors to record information about muscle activity when using different pointing devices. The results of the study will be used to determine the benefits and feasibility of integrating these components into the future en route environment.

**Experimental Procedures:**

Twenty-four en route Certified Professional Controllers (CPCs) from Level 11 and 12 facilities will participate in the simulation.[4] Four participants will arrive at the lab at a time. They will spend 8 days at the lab over a 2-week period. They will travel in on a Monday and travel out on Friday of the following week. At the start of the simulation, the participants will be randomly assigned to work as either an R-side or D-side controller for conditions that require R-side/D-side teams. Each participant will remain in the assigned position for all conditions that require teams. Other experimental conditions will require that each participant work as an R-side alone.

The participants will work from about 8:00 AM to about 4:30 PM every day with a lunch break and at least two rest breaks. The first morning will consist of an initial briefing to review project objectives and participant rights and responsibilities. It will include initial familiarization training on the simulated airspace, the system, and the procedures. The participants will then go the laboratory to begin hands-on training on the first of the three simulation components. They will complete practice scenarios prior to completing the test scenarios. All scenarios will be about 45-minute in duration.

During designated scenarios in the conflict probe and display evaluation components of the simulation, the R-side participants will wear a head-mounted oculometer to record eye movement data via infrared technology. The exposure to infrared illumination while wearing the oculometer is less than 4% of the intensity of that experienced when outside on a sunny day.

During designated scenarios in the pointing device evaluation, the participants will wear wireless electromyographic (EMG) recording sensors to obtain data about arm and wrist movement and muscle activity when using each device. The participants will wear short- sleeve shirts to allow access to the upper arm and upper back/neck area. Sensors will be applied to the skin using hypoallergenic gel, adhesives, and Velcro straps. Before attaching the sensors, we will wipe the skin with alcohol pads and/or an abrasive skin cleanser to remove oils to obtain the clearest possible signals. For some participants, we may also need to trim hair or shave small areas of skin at the location where the sensors will be applied.

The participants will provide workload ratings when prompted at designated intervals throughout each scenario. An automated data collection system will record system operations and generate a set of standard Air Traffic Control (ATC) simulation measures, including safety, capacity, efficiency, and communications. After each

[4] This sentence was modified in the form provided to the retired controllers.

scenario, the participants will complete questionnaires to report their overall workload, situation awareness, and performance and to rate various aspects of the test condition. The simulation will be audio and video recorded.

After the participants have completed each of the simulation components, they will gather for a final debriefing session to provide final comments and feedback.

**Anonymity and Confidentiality:**
My participation in this simulation is strictly confidential. Any information I provide will remain anonymous: no individual names or identities will be associated with the data or released in any reports.

**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 effectiveness of potential ATC tools and workstation configurations. My data will help the FAA to determine the benefits and feasibility of these modifications in this environment.

**Participant Responsibilities:**

I am aware that to participate in this study I must be a certified professional controller who is qualified at my facility and holds a current medical certificate. I must also have normal or corrected-to-normal (20/20) vision and do not wear bifocals, trifocals, or hard-contact lenses that are incompatible with the eye-tracking device used in this simulation. I will control traffic and answer the questions asked during the study to the best of my abilities. I will not discuss the content of the experiment with anyone until the study is completed.

**Participant Assurances:**

I understand that my participation in this study is completely 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, I will be informed. 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 Dr. 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 Dr. Zingale at (609) 485-8629.

**Discomfort and Risks:**

I understand that I will not be exposed to any foreseeable risks or intrusive measurement techniques. The only anticipated discomfort may be some skin redness at the site of the EMG sensor placement or some discomfort from the oculometer head mount. I agree to immediately report any injury or suspected adverse effect to Dr. Carolina Zingale at (609) 485-8629.

**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:__________

Appendix B: Background Questionnaire

**Background Questionnaire**

Instructions:

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.

Demographic Information and Experience

| | | |
|---|---|---|
| 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. Rate your current skill as a CPC. | Not Skilled | ①②③④⑤ ⑥⑦⑧⑨⑩ | Extremely Skilled |

| | | | |
|---|---|---|---|
| 9. Rate your level of motivation to participate in this study. | Not Motivated | ①②③④⑤ ⑥⑦⑧⑨⑩ | Extremely Motivated |

# Appendix C: Post-Scenario Questionnaire

## Post-Scenario Questionnaire

Instructions:
Answer the following questions based upon your experience in the scenario just completed. Fill in one circle to indicate your response to each item.

**Part 1 – Overall Performance, Workload, Situation Awareness, and Simulation Ratings**

| | | | |
|---|---|---|---|
| 1. Rate the **overall difficulty** of this scenario. | Extremely Difficult | ①②③④⑤⑥⑦⑧⑨⑩ | Extremely Easy |
| 2. Rate your **overall level of ATC performance**. | Poor | ①②③④⑤⑥⑦⑧⑨⑩ | Excellent |
| 3. Rate your **overall workload**. | Poor | ①②③④⑤⑥⑦⑧⑨⑩ | Excellent |
| 4. Rate your **workload due to communications** with pilots. | Poor | ①②③④⑤⑥⑦⑧⑨⑩ | Excellent |
| 5. Rate your **overall level of situation awareness**. | Poor | ①②③④⑤⑥⑦⑧⑨⑩ | Excellent |
| 6. Rate your **situation awareness for *current* aircraft locations**. | Poor | ①②③④⑤⑥⑦⑧⑨⑩ | Excellent |
| 7. Rate your **situation awareness for *projected* aircraft locations.** | Poor | ①②③④⑤⑥⑦⑧⑨⑩ | Excellent |
| 8. Rate your **situation awareness for potential aircraft loss-of-separation**. | Poor | ①②③④⑤⑥⑦⑧⑨⑩ | Excellent |
| 9. Rate your **situation awareness for potential handoff/airspace violations**. | Poor | ①②③④⑤⑥⑦⑧⑨⑩ | Excellent |
| 10. Rate the **performance of the simulation pilots** in terms of their responding to control instructions and providing callbacks. | Poor | ①②③④⑤⑥⑦⑧⑨⑩ | Excellent |

11. What aspects of this scenario were easiest to work with? Why?

12. What aspects of this scenario were hardest to work with? Why?

13. Do you have any additional comments or clarifications about your experience in this scenario?

**Part 2: Conflict Probe Location and Format**

These questions pertain to your experience working with the conflict probe presentation in the scenario just completed. Fill in one circle to indicate your response to each item.

| | | | |
|---|---|---|---|
| CP1: Rate the overall effectiveness of the **location** of the conflict probe. | Poor | ①②③④⑤⑥⑦⑧⑨⑩ | Excellent |
| CP2: Rate the overall **accuracy** of the probe information. | Poor | ①②③④⑤⑥⑦⑧⑨⑩ | Excellent |
| CP3: Rate the overall effectiveness of the conflict probe **format** (e.g., colors, icons, flashing) in alerting you to a potential conflict. | Poor | ①②③④⑤⑥⑦⑧⑨⑩ | Excellent |

C4. Please provide any additional information about your experience working with the conflict probe in this scenario.

**Part 3: Display Monitors**

These questions pertain to your experience working with the display monitor provided in the scenario just completed. Fill in one circle to indicate the extent to which you agree with each statement.

| | | | |
|---|---|---|---|
| D1. The displays were uncluttered. | Completely Disagree | ①②③④⑤⑥⑦⑧⑨⑩ | Completely Agree |
| D2. I could find the information I needed quickly. | Completely Disagree | ①②③④⑤⑥⑦⑧⑨⑩ | Completely Agree |
| D3. The text was easy to read (contrast, size, etc). | Completely Disagree | ①②③④⑤⑥⑦⑧⑨⑩ | Completely Agree |
| D4. The graphics were legible and easy to interpret. | Completely Disagree | ①②③④⑤⑥⑦⑧⑨⑩ | Completely Agree |
| D5. The display was in an ideal physical position. | Completely Disagree | ①②③④⑤⑥⑦⑧⑨⑩ | Completely Agree |
| D6. The display was an appropriate size. | Completely Disagree | ①②③④⑤⑥⑦⑧⑨⑩ | Completely Agree |
| D7. The display was an appropriate shape. | Completely Disagree | ①②③④⑤⑥⑦⑧⑨⑩ | Completely Agree |

D8. Please provide any additional information about your experience using the display provided in this scenario.

## PSQ Comments

## Conflict Probe Evaluation

| Conflict Probe Evaluation – Post Scenario Questionnaire Comments<br>(R-side controllers working alone)<br>***PSQ Question:*** *What aspects of this scenario were easiest to work with? Why?* | | |
|---|---|---|
| | **CONDITION** | |
| **BASELINE** | **BPLS** | **PLPL** |
| Ability to vector out of situations. | Frequencies | Frequencies |
| Not a lot of distractions from colors and unnecessary alerts. Familiarity with problem. More like what I am use to. No distractions from D-side. | Conflict alert list on R-side - Did not have to go to D-side to examine conflict. | Conflict probe on R-side. |
| | | Conflict alert indicator right over the data block. Good scale to work on. |
| | Being able to highlight data block with only one click. | Conflict alert probe (make traffic search and traffic resolution for controller much easier when associated with data block). |
| Less clutter with fewer distractions. Work at maximum range to observe incoming traffic and optimum font size to reduce clutter. | The scope (allowed maximum range), able to expand visual range and keep data blocks apart. | Being familiar with the scenario. |
| Familiarity of scenario. | The redundancy. | Drop down menus for route. Probes on R-side – easily recognized. Macros – group function. |
| Drop down menus for routing. Macro for altitude and routing. Multi entry inputs. Able to accomplish more simultaneously. | R-side conflict probe. Intuitive and easy to use. Still need to gray out examined alerts. | |
| | Route menu, drop down menu. Accessibility | |

**Conflict Probe Evaluation – Post Scenario Questionnaire Comments**
**(R-side controllers working alone)**

***PSQ Question:*** *What aspects of this scenario were hardest to work with? Why?*

| CONDITION | | |
|---|---|---|
| **BASELINE** | **BPLS** | **PLPL** |
| Looking away to "D" side to address conflicts, then returning to the radar display- too much "off glass" time. | Miss the drag and drop feature. Handoff not being accepted or handoff complete, switched, then starts flashing back. | The time it takes to do the CP takes away from vector time. |
| Working "one hole" and trying to acknowledge D side probe. | The conflict probe menu- acknowledgment. | Overlapping data blocks- harder to pull them apart. |
| No drag and drop data blocks. | Conflict alert list contained too much information, not usable to look for real conflictions. | No drag and drop data block – drag and drop allows exact placement. |
| No conflict probe meant you had to be more alert. D-side conflict probe useless because of the business of the problem. | Conflict alert probe not readily available (box), or on D-side which when busy was ignored in favor of keeping eyes on scope. | Conflict alert box too full with unneeded alerts. Unnecessary alerts on arrival aircraft. |
| D-side data virtually useless to R-side when busy. Small GPD and too many data blocks make this position useless. | Conflict probe is too far out making it too complicated to use. | The sensitivity of eye tracker (even though normal range of motion is tolerated you are cognizant that slight alterations in optical device could cause problems with data. |
| Looking to the ACL for conflict probe information. | | The conflict probe projects too far out causing me to ignore most alerts. |
| I was getting really busy at the end and it was annoying to have to look at the probes on the D-side. | | Recognizing which probe alerts have been acknowledged because w/o resolving alert, it remains. |
| Data block overlap, congestion. Pilot read back- self expl. | | |

**Conflict Probe Evaluation – Post Scenario Questionnaire Comments**
**(R-side controllers working alone)**

***PSQ Question:*** *Do you have any additional comments or clarifications about your experience in this scenario?*

| CONDITION | | |
|---|---|---|
| **BASELINE** | **BPLS** | **PLPL** |
| Reviewing the conflicts on the "D" side- similar complaint- once we have reviewed, no indication that the task is completed. When several handoffs are flashing at us from the same location, the FDB's are all on top of each other, this is just a programming problem.<br><br>Went to small font size (#2) which made data block overlap better, but left the cursor the same size causing problems when trying to pick a route. Smaller cursor an option, but awful small.<br><br>Two things could be useful to controller in this situation:<br><br>1. Data blocks automatically seek non-overlap position when in close proximity to other data blocks.<br><br>2. Conflict alert probe in association with data block. | EQF4334- code 2431- HO never taken. AAL7592- code 2108- handed off to sector 53 then started flashing back at 08.<br><br>Conflict probe for crossing traffic is helpful, but conflicts identified on same route or opposite direction route that conflicts with all are too numerous.<br><br>Much better without a D-side.<br><br>It would have been better to have conflict alert probe associated with the data blocks in conflict.<br><br>I did not use conflict probe list.<br><br>I forgot to mention earlier that it's really cool to be able to click in the altitude field and see green, yellow and red altitudes indicating confliction free altitudes. It's a neat concept to see the same feature in the speed field, but it's not as accurate.<br><br>Yes- I am tired! | Am curious, not so much on the 2K display, but on the 30-inch display, the LDB's in the AOI area will be a big distraction.<br><br>2K display much better.<br><br>Bright green route/alert line too bright and distracting. Nice being able to highlight a data block with a single click. Used a smaller font which helped with data block overlap, but a little harder on an old man's eyes.<br><br>If it isn't already in the field, incorporate conflict alert probe being associated with data block as quickly as possible.<br><br>Six min probe info is easier to work with and trust. |

**Conflict Probe Evaluation – Post Scenario Questionnaire Comments**
**(Team Configuration: R-side)**

***PSQ Question:*** *What aspects of this scenario were easiest to work with? Why?*

| CONDITION | | | |
|---|---|---|---|
| **BASELINE** | **FLUR** | **PLPL** | **6 MIN** |
| Knowledge of adjacent sector frequencies. | Drop and drag data blocks. | The only easy thing is knowing the frequencies. | Drag and drop is back! |
| Making macros increased efficiency. | Drag and drop data blocks – allows exact positioning. | Rerouting aircraft with the drop down fix menu is very helpful as well as big time saver. | Drop and drag data blocks. |
| Pretty much the same as in the field today. | Like being able to slew on data block and move it. Like being able to get a route readout by clicking on the data block. | Conflict probe on R-side – Do not have to look away from R-glass. | Drag and drop data blocks – can position exactly where you want them. Conflict display right on the R-side – do not have to work of the D-side. |
| Not a lot of clutter on the screen. Familiarity with same problem. | Auto hand off feature (always on), conflict alert probe on (see potential traffic). | We had seen the same scenario over and over. Conflict probe location was useful. | The 6 min/6 mile was just right for alerting the R-side. |
| Keyboard vs. macros (unless using single entry for multiple aircraft it is still easier and faster to use keyboard for computer entries, routes, etc.) | Familiar with situation. | Auto hand off feature never inhibited, ++ feature for detecting traffic and preplanning actions. | "Adjustable" GPD and ACL screen on D-side made the position available and useable. (The GPD could be increased in range and placed side by side with ACL). |
| Same scenario every time. | Routing drop down menus. Accessibility. | Repetition we have seen the same situations many times. | Repetition |
| No "rubber band" data blocks is a good thing. | | R-side conflict probe is a fantastic idea. It keeps controller focused on traffic. | 6 + 6 R-side conflict probe is awesome. I could glance at the D-side to view probes further in the future. |
| Drop down menu for routing because it is easily accessible. | | Drop down menus because of accessibility on the glass. | Drop down menus due to being on the glass. Click and drag data blocks able to position without keyboard. |

**Conflict Probe Evaluation – Post Scenario Questionnaire Comments**
**(Team Configuration: R-side)**

***PSQ Question:*** *What aspects of this scenario were hardest to work with? Why?*

| CONDITION | | | |
|---|---|---|---|
| **BASELINE** | **FLUR** | **PLPL** | **6 MIN** |
| Re-locating data blocks to prevent overlap. | The overlapping of data blocks. | Occasionally, D side would advise me he just took a handoff, this broke my concentration. | Near the end of the problem when traffic is near peak, you are constantly moving FDB's to prevent overlap. |
| Volume of landing traffic. You need to ensure they are spaced appropriately. | Amount of data blocks. | Conflict probe can become distracting making it difficult to plan moves when concentration broken. | Sometimes drop and drag data blocks are too sensitive to cursor position which results in selecting diff. menu. |
| No drag and drop data blocks. No conflict probe on R-side. | The way the "leader line" associates with the data block was somewhat disorienting. | No drag and drop data block. | Pilots seemed to miss many alt + freq. |
| Same conflict probes were showing yellow, yet the aircraft were getting uncomfortably close. | The bright green line for reroutes and conflicts was a distraction. The conflict alert red box is also distracting. Need to be able to suppress the bright red conflict. | Data block overlap became a problem at times (as intended). | Leader lines not associated with data block intuitively. |
| D-side GPD (Too many data blocks in D-side GPD for making it useful.) | Conflict alert probe on for landing traffic (information obtained low priority). Conflict alert on ACFT out of my airspace not necessary. | Data block overlap (makes it more difficult to see/separate traffic) | Bright green light too much. Route readout should only display what is asked for , not the entire route. |
| Conflict probe only on the ACL. | Conflict probe projects too far out and the rubber band data block leader line function when trying to highlight the data block. It makes highlighting the data block harder because two clicks are required. | Too many conflict probe alerts too far out. | ACL pops up on radar screen every time controller uses conflict alert probe. Useless feature – you already have conflict probe associated with data block – not necessary to have box pop-up. |
| Lack of R-side conflict probe is bad. | Volume of aircraft and answering calls. | The probe was less accurate and gave me information way too far in advance. I need more imminent alerts on R-side. | It would be hard if unfamiliar with the scenario. |
| | | Volume of traffic was significant. | Sheer volume. A/c entering sector wrong altitude for direction of flight. Personally, I don't like Genera (Generic) airspace. Pick a real sector. |

**Conflict Probe Evaluation – Post Scenario Questionnaire Comments**
**(Team Configuration: R-side)**

***PSQ Question:*** *Do you have any additional comments or clarifications about your experience in this scenario?*

| CONDITION | | | |
|---|---|---|---|
| **BASELINE** | **FLUR** | **PLPL** | **6 MIN** |
| With no conflict probe and a D-side actively using his conflict probe, the tendency is to relax more and not concentrate on looking for conflictions, but to separate data blocks and watch for handoffs.<br><br>#12 Either D-side should have capability to expand GPD or when busy D-side should shut down GPD, use ACL and correlate traffic using R-side scope. | Previously documented unrealistic conflict indications. Ex: 30 minutes apart faster in front.<br><br>Too many key clicks to highlight and un-highlight a data block.<br><br>It appeared that sometimes were (?) and sometimes moved (keeps controller from becoming too complacent and routine).<br><br>Pilots did much better this scenario. | D side could assist more with taking and initiating handoffs.<br><br>Pilot read-backs have to be listened to very carefully.<br><br>Aircraft in the vector when the problem begins should not have to check on frequency.<br><br>Vary the times ACFT enter problem, if not the problem itself. To counter balance controller "familiarity" with problem, make them do P/O's.<br><br>Yes, the 6 min six mile probe is much easier to use and pay attention to. | The ability to move the data block anywhere you want is nice, but having to triple click a data block to highlight and un-highlight, plus too many inadvertent responses make it not really worth it.<br><br>#11 This scenario for D-side much more useable than baseline D-side (which was unusable for R-side) and barely usable for D-side (worked best if you shut down GPD and used ACL only) – using R-side radar for correlation. Useful; interactive data blocks. Adjustable GPD and ACL screen for D-side, removal of conflict alert probe from R-screen.<br><br>I avoided many potential conflicts by already seeing the problem and fixing them before the 6 min. |

**Conflict Probe Evaluation – Post Scenario Questionnaire Comments**
**(Team Configuration: D-side)**

***PSQ Question:*** *What aspects of this scenario were easiest to work with? Why?*

| CONDITION | | | |
|---|---|---|---|
| **BASELINE** | **FLUR** | **PLPL** | **6 MIN** |
| Conflict alert on RA side was good. | Conflict probe really worked. | Conflict alert was right on. | Conflict alerts were correct – yellow vs. red was correct and good info. |
| Moving data blocks. Direct interaction with R-side. | Keeping up w/ data blocks. Familiarity. | Everything. 4 days of repetition | Moving data blocks. |
| Managing ACL. | No reroutes! All a/c flew filed routes Repetition. | Traffic flow | Trackball and monitor size. |
| Selecting GPD/ACL. | Display / because of prefsets and adjustability. | | Large interactive screens allow simultaneous viewing of ACL and GPD. Prefsets!!! |
| Mouse logic on trackball / The only possible way to separate data block overlap. | All scenarios – D-side dwelling on a/c in ACL highlighting call sign in data block -> helpful characteristics. | | Radar data on D-side GPD – very easy to use – to the point where I don't even look at the radar scope. |
| Familiar D-side. | 30-inch monitor is sweet. | | |
| ERAM, familiar with. | | | |

**Conflict Probe Evaluation – Post Scenario Questionnaire Comments**
**(Team Configuration: D-side)**

***PSQ Question:*** *What aspects of this scenario were hardest to work with? Why?*

| CONDITION | | | |
|---|---|---|---|
| **BASELINE** | **FLUR** | **PLPL** | **6 MIN** |
| Too many conflict alerts- too far out from actual time.<br><br>ACL. Bottom line kept dropping off scope.<br><br>Aircraft saturation.<br><br>Data block congestion, conflict probe resolution, monitor size way too small.<br><br>Size of display and GPD / Can't view all FP info, too much clutter. | Too many red alerts too far away- not urgent!<br><br>Separating data block. Saturation.<br><br>None (Maybe data block o-lap hard to keep up with).<br><br>The need to recenter display after every alert check / Increases workload.<br><br>6 mile 6 minute conflict alert -> five mile would be much better.<br><br>Alerts, GPD seems easier to me. | Too many CA's too far out and on arrivals which would conflict "down low."<br><br>None. Occasional tracking device failure.<br><br>Size and brightness of display (not adjustable) too small, too bright.<br><br>Alerts appear too early. | GPD. When viewing the conflict w/ route displayed, the screen size didn't revert to the previous scale when cleared.<br><br>Re-centering the GPD after displaying a conflict.<br><br>Data block saturation.<br><br>Still, inability to suppress alerts after evaluation. Don't know which have already been checked.<br><br>The D-side functions only affect the D-side makes it difficult to assist R-side. |

**Conflict Probe Evaluation – Post Scenario Questionnaire Comments**
**(Team Configuration: D-side)**

***PSQ Question:*** *Do you have any additional comments or clarifications about your experience in this scenario?*

| CONDITION | | | |
|---|---|---|---|
| **BASELINE** | **FLUR** | **PLPL** | **6 MIN** |
| No.<br><br>I'm beginning to know this problem by heart<br><br>Conflict probe "evaluation box" nuisance. Should be displayed after resolution/review or not at all. Congestion was so bad couldn't evaluate efficiently. Must create filtering capability for approach airspace conflictions.<br><br>Need to be able to scroll through all flight plans in the ACL. GPD should re-center automatically after each conflict probe check. (Also includes range return.)<br><br>This conflict probe not desirable.<br><br>No. | Red alerts need to be imminent- not 15-20 minutes ahead, it loses its urgency if too many are red.<br><br>Change times on some aircraft. Allowing a/c to fly filed route (no r-rts) increased complexity, more fun to resolve conflict potentials.<br><br>In general, I believe that not being able to do anything else without first answering the conflict evaluation questions detracts from accuracy of said evaluations. | No.<br><br>Change 2/3 aircraft times to regulate traffic flows differently.<br><br>Addition of 1 or 2 different aircraft on different crossing routes would increase difficulty substantially. | Too much information on far away conflicts at end of problem.<br><br>Still saw con-probe not reflecting traffic @ 5.41 miles. Easy scenario with tools supplied, but re-center and rearranging is not necessary<br><br>Trackball logic for moving data blocks is great, except occasionally center-click will not release (most clear).<br><br>Yes, because of my answer to question 12 it made it hard to focus on problem. |

# PSQ Comments

## Display Evaluation

| Display Evaluation – Post Scenario Questionnaire Comments (R-side)<br>***PSQ Question:*** *Please provide any additional information about your experience using the display provided in this scenario.* | | | |
|---|---|---|---|
| **CONDITION** | | | |
| **2K1K** | **2K30** | **303R** | **303S** |
| The ability to configure the scope to your prefs is helpful.<br><br>Probed route display is too bright.<br><br>Anything that can be done to work on a smaller range is better to avoid data block overlap.<br><br>2k is easier to use than the 30". | Working a long sector did not allow me to work on a good range. | Needed to change range in this config.<br><br>Display shape did not work well with sector shape.<br><br>The display shape, for this sector was appropriate. However, the orientation was incorrect. Should be in a portrait mode for this sector.<br><br>The display was uncluttered except for data blocks. The display had too much unused space when more space was needed to enlarge the map and work on a better range. If the display could have been turned vertical, it would have been perfect.<br><br>With practice this screen would be more comfortable to use. | The eye tracker hardware tends to settle over time and pushes forward. Sometimes effects vision.<br><br>Would be nice to rotate display to portrait mode for this sector.<br><br>The overall display had lots of wasted space on the side, but the vertical distance forced the map to be on a compact range. Went to a smaller font to prevent data block overlap.<br><br>I believe with practice there would be no loss of comfort or usability.<br><br>The 30 inch gives too much mapping information east to west for a sector that is narrow and runs north to south. |

| **Display Evaluation – Post Scenario Questionnaire Comments (D-side)** <br> ***PSQ Question:*** *Please provide any additional information about your experience using the display provided in this scenario.* | | | |
|---|---|---|---|
| **CONDITION** | | | |
| **2K1K** | **2K30** | **303R** | **303S** |
| Too much information on display- it spilled out of the bottom and I couldn't keep track of the "big picture" as to how many acft were actually inbound to the sector | Large display a definite plus. Easier to read/scan lists. Easier on my eyes. | Monitor easy to see- easy to set up ACL and GPD next to each other. When examining a conflict, the 'overlaid' screen changes size/range seemingly at random- requiring reset at every examination! | Easy to read. No problems keeping data blocks clear and readable in the GPD- although it was too bright. |
| This display was very easy to see without straining. | This display was almost too big. | Much less clutter than last time. | Cannot watch FDB's on GPD due to constant resizing. |
| URET functionality was inconsistent at best. | Having to recenter display GPD after conflict probe was very time consuming. Not necessary suggest "map range" change temp or "map center" temp until probe is reviewed and/or resolved. | Allow display to rotate (portrait vs. landscape). | Like the big monitors! |
| Size unacceptable for GDP and ACL usage! Could not acquire ACL FP without minimizing GDP. (IMO) bad idea, way too cluttered. Again, temp alt do not generate con-probe alerts. | Need complete contrast and brightness controls (prefsets would be ideal). | So much more user friendly using prefsets and advanced functionality. Still need to consider the "recenter" function inclusive after conflict probe is resolved/reviewed. | Larger display is very good, capable of simultaneous ACL and GPD. |
| Too much clutter, spend unacceptable amount of time simply trying to keep data blocks from overlapping. Unable to control brightness and contrast of GDP. | | Prefsets and advanced D-side functionality provide for excellent working conditions. | |
| Ideal setup for D-side screen would be 30" monitor with 1k ACL. Larger, more functional GPD with 6 minute conflict alert. Ability to move D-side screen is necessary. | | Inability to control data blocks from D-side without typing FWD is time consuming. | |

Appendix D: Exit Questionnaire

## Exit Questionnaire

### Part 1 – Simulation Realism and Research Apparatus Ratings

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

| | | | |
|---|---|---|---|
| 1. Rate the **realism of the generic airspace** compared to actual ATC operations. | 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 online workload rating technique interfere** with your ATC performance? | Not At All | ①②③④⑤⑥⑦⑧⑨⑩ | A Great Deal |
| 6. How effective was the training provided? | Not At All Effective | ①②③④⑤⑥⑦⑧⑨⑩ | Extremely Effective |
| 7. Answer only if you wore the oculometer: To what extent did the oculometer interfere with your ATC performance? | Not At All | ①②③④⑤⑥⑦⑧⑨⑩ | A Great Deal |
| 8. Answer only if you wore the EMG sensors: To what extent did the EMG monitoring equipment interfere with your ATC performance? | Not At All | ①②③④⑤⑥⑦⑧⑨⑩ | A Great Deal |

Exit1. Please include any additional comments about the simulation that you would like us to know about.

**Part 2 – Conflict Probe Location and Format**

These questions pertain to your experience working with the conflict probe in different locations and in different formats in the simulation. Please fill in one circle to indicate your response to each item.

| Conflict Probe Format | Much better with color-coded alerts | Somewhat better with color-coded alerts | No Difference between color-coded alerts and flashing alerts | Somewhat better with flashing alerts | Much better with flashing alerts |
|---|---|---|---|---|---|
| 9. Detecting potential conflicts | ① | ② | ③ | ④ | ⑤ |
| 10. Determining when a potential conflict was expected to occur | ① | ② | ③ | ④ | ⑤ |
| **Conflict Probe Location** | **Much better on the D-side only** | **Somewhat better on the D-side only** | **No Difference whether on the D-side only or on both the R-side and D-side** | **Somewhat better when on both the R-side and D-side** | **Much better when on both the R-side and D-side** |
| 11. Detecting potential conflicts | ① | ② | ③ | ④ | ⑤ |
| 12. Determining when a potential conflict was expected to occur | ① | ② | ③ | ④ | ⑤ |
| **Conflict Probe Layout** | **Much better with the List Only on the R-side** | **Somewhat better with the List Only on the R-side** | **No Difference between the List Only on the R-side and the List + Data Block indicators** | **Somewhat better with the List + Data Block Indicators on the R-side** | **Much better with the List + Data Block Indicators on the R-side** |
| 13. Detecting potential conflicts | ① | ② | ③ | ④ | ⑤ |
| 14. Determining when a potential conflict was expected to occur | ① | ② | ③ | ④ | ⑤ |

Exit2. Please include any additional comments on the location of the conflict probe and the different types of conflict probe formats and layouts you experienced in this simulation.

**Part 3 – Display Monitors**

These questions pertain to your use of different display monitors in the simulation. Please fill in one circle to indicate your response.

| | Much Better with a 30" monitor | Somewhat Better with a 30" monitor | No Difference Between 30" monitor and the standard display | Somewhat Better with a standard display | Much Better with a standard display |
|---|---|---|---|---|---|
| 17. Locating information on the display | ① | ② | ③ | ④ | ⑤ |
| 18. Scanning traffic effectively | ① | ② | ③ | ④ | ⑤ |
| 19. Avoiding or resolving potential conflicts | ① | ② | ③ | ④ | ⑤ |
| 20. Maintaining situation awareness | ① | ② | ③ | ④ | ⑤ |
| 21. Managing traffic efficiently | ① | ② | ③ | ④ | ⑤ |

Exit3. Please include any additional comments on your experience using the different types of display monitors.

# Exit Questionnaire Comments

## Conflict Probe Evaluation – Exit Questionnaire

*Please include any additional comments on the location of the conflict probe and the different types of conflict probe formats and layouts you experienced in this simulation.*

**R-side participant responses:**

- When CP was on the radar or on the "D" side and the radar controller had to manipulate the list it was not user friendly. Once you opened the conflict to analyze it and went to open it again to close out the graphics on the scope, the line was not easily recognized. The box that (or line) we are working should either remain highlighted or an edge around it so we can easily pick it and close out the graphics on the scope. On the "D" side- same thing, except you open an alert to analyze, complete the prompts and close it out. Then there is no indication that you have viewed the conflict and as time goes by, when you open it again, you know that you have completed this for the prompt is not available. Some symbology should be used so that we don't repeat actions. I felt the easiest format was on the "O" line, just a single digit. The [boxed A] was too bright, especially with the flashing. Most the time the confliction has been noted and this serves as a reminder.
- The red or yellow numbers above the data block gave me a good indication of what to expect as far as future conflictions. The list on the R side was very distracting. The flashing [boxed A] was also a distraction.
- I did not look at the conflict list on the R-side much, but I think this is likely because I already knew where all the conflicts occur in the scenario.
- When you start getting busy it is very helpful to have the conflict probe on the R-side.
- Only want probe in the data block, not in the list on the R-side. The 6 mile/6 minute was the best format. Not as many unnecessary alerts or alerts that were happening too far in advance to be useful. The conflict alert list on the R-side was not helpful because you are concentrating on your traffic, especially when busy. The flashing alert was distracting. The brightness of the route for both route readouts and conflict probe was too bright.
- [Re: Conflict Probe Location] - useful from D-side when not using Baseline format (Baseline format does not allow to scroll ACL list or make GPD larger. [Re: Conflict Probe Layout] - Preferred to have conflict probe on data block only - no list popping up on R-screen. [Exit2 Additional Comments] - Working the final scenario was best for conflict probe. One the D-side ACL and GPD were allowed to be expanded giving better views of traffic (GPD) and data (ACL). From the R-side having the conflict probe associated with data block was a time saver and less distracting. However, having the conflict probe box appear every time the controller initiated a conflict probe search was useless and distracting.
- The conflict probe layout list on R-side is cumbersome and the lack of changing a color or denoting entry being acknowledged makes the list useless for what I was trying to accomplish.
- I liked the numbers above the R-side data block for the alerts.
- Reduced conflict probe parameters are a plus because it reduces the unnecessary alerts that are displayed on the R-side. The D-side still has need to have the probe further away for planning purposes and that way better assist the R-side. Ideally, all colors need to mute once the specific alert has been selected so actions aren't needlessly duplicated.
- Having the potential conflict light up on the radar scope was pretty impressive and could be a great asset in the field. The color coding was definitely great because it quickly let you see what level of action was needed if a conflict existed.

**D-side participant responses:**

- Red blocks w/ white "A" in field 0 wasn't effective for me on "R". Red and yellow numbers in field 0 worked well with me on "R". Last configuration run w/ only red numbers wasn't effective for me- the red numbers lost their effectiveness with alerts more than 7-8 minutes away. My favorite and most useful display had red and yellow numbers in field 0 in the data block- easy to read from the RA ("D") side.
- I didn't find the list to be useful on either the "R" or the "D" side. As long as I know that there is a potential conflict, the graphic display is more than adequate. I did like the flashing indicator better than today's flashing data block. It gets your attention without being distracting.
- Didn't like the flashing probe on the R-side.
- Probe list on R-side should be selectable. Easy to run without, but can be useful. Appears to clutter monitor. Too much information. <6 min separation might be buffered to "7" miles. Some scenarios actually did not light with <6 nmi. Add filtering capability to exclude "approach airspace" once separation is ensured in en route or "my" airspace. Next sector probing should not display. Allow probing to adjacent facility airspace when vectoring is used.
- I cannot honestly declare that the probe location and/or layout being D-side and/or R-side makes a difference since I didn't have the opportunity to work the R-side except during musculoskeletal testing scenarios. 6 minute/6 mile format is a far superior parameter for ATC purposes. Probing below or above the sector altitude stratum and/or the set altitude filter limits should be eliminated. It's a distraction. There needs to be a way to suppress alerts after having evaluated them, therefore, eliminating the secondary or repetitive acknowledgments of the same alert.

**Display Evaluation – Exit Questionnaire**

Display Monitors*:*

*Please include any additional comments on your experience using the different types of display monitors.*

**R-side participant responses:**

- For the type of sector we were working (tall) the standard display worked best. On the 30 inch the adjustments made the sector feel compressed, which enhanced the overload of FDB. However, this could probably be said if the sector was long- there the 30 inch would be better.
- The example sector did not fit well on the 30 inch monitor. I had to change the range to accommodate the shape. Maybe an option to rotate the screen, similar to the ERID's display could be looked at.
- I think that the 30" display would be superior, for this sector, if it could be rotated into the portrait mode.
- The 30" had too much wasted space using a sector like we used. It would be useful on a sector that is more horizontal than vertical. Had to work on a larger range that caused too much data block overlap. Tried using a smaller font and cursor than used to, but that created problems seeing the text, especially for the D-side.
- Using the 30" monitor on the R-side was more difficult. Data and displays were more "compacted" making it much more difficult to "see" and separate traffic.
- I find the standard display easier and more comfortable to work with. However, with practice, believe I would become more comfortable with the 30" monitor.
- Working a tall sector like Genera, it made no difference to me which monitor I used. If I was working an East/West sector, I probably would strongly prefer the 30" widescreen monitor.
- The 30" was nice to put all data needed on one screen, but seemed to discourage radar teamwork.
- My responses are from the R-side perspective only. I firmly believe that the 30" monitor is a far superior display for the D-side.

**D-side participant responses:**

- I felt that the larger displays gave me more "room" to see the traffic picture. I'm not sure how much more spread out the range or scale was (if it was at all)- but it seemed less cluttered to me.
- 30" monitor displays way too much external area, especially with a north/south oriented sector.
- Enable portrait or landscape orientation.
- Use of 30" over smaller display is like night and day. Larger display allows for simultaneous use of ACL and GPD, while reducing clutter on the GPD. Also, prefsets and adjustability are an enormous asset.
- I prefer 30" screen on the D-side, but it does greatly increase the amount of things you have to include in your D-side scan- increasing workload unnecessarily making me unable to assist the R-side as much as I could.