Situation Awareness in Air Traffic Control: Enhanced Displays for Advanced Operations

Mica Endsley, Ph.D., SA Technologies
Randy L. Sollenberger, Ph.D., ACT-530
Akiko Nakata, ACT-530
Earl S. Stein, Ph.D., ACT-530

January 2000

DOT/FAA/CT-TN00/01

U.S. Department of Transportation
Federal Aviation Administration

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

NOTICE

Technical Report Documentation Page

1. Report No. DOT/FAA/CT-TN00/01	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Situation Awareness in Air Traffic Control: Enhanced Displays for Advanced Operations		**5. Report Date** January 2000 **6. Performing Organization Code** ACT-530
7. Author(s) Mica Endsley, Ph.D., SA Technologies, Randy Sollenberger, Ph.D., Akiko Nakata, and Earl S. Stein, Ph.D., ACT-530		**8. Performing Organization Report No.** DOT/FAA/CT-TN00/01
9. Performing Organization Name and Address Federal Aviation Administration William J. Hughes Technical Center Atlantic City International Airport, NJ 08405		**10. Work Unit No. (TRAIS)** **11. Contract or Grant No.**
12. Sponsoring Agency Name and Address Federal Aviation Administration Human Factors Division 800 Independence Ave., S.W. Washington, DC 20591		**13. Type of Report and Period Covered** Technical Note **14. Sponsoring Agency Code** AAR-100
15. Supplementary Notes		

16. Abstract

Future changes in the National Airspace System indicate a self-separation operational concept. This study examined the Air Traffic Control Specialist's ability to maintain situation awareness and provide needed monitoring and separation functions under this concept. The study also provides an empirical evaluation of the effects of an enhanced display concept as a window on the existing air traffic control radar display. This window furnished the targeted altitude or heading of aircraft in a transitionary state. This information simulated that provided by a datalink from the aircraft flight management system in a future air traffic operation. The introduction of the enhanced display appeared to have some utility in aiding controllers in dealing with air traffic operating under self-separation. The enhanced display resulted in improvements in some air traffic control performance and situation awareness measures. Results suggest that further improvements in the enhanced display may better integrate it with the controller radar display. In addition, researchers recommend methods for measuring controller situation awareness in future air traffic control research.

17. Key Words Situation Awareness Free Flight Shared Separation Advanced Displays		**18. Distribution Statement** This document is available to the public through the National Technical Information Service, Springfield, Virginia, 22161	
19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 55	**22. Price**

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

Table of Contents

Appendixes

A - Subject Matter Expert Subjective Rating Form
B - Situation Awareness Global Assessment Technique Queries
C - 10-Dimensional Situational Awareness Rating Technique
D - Air Traffic Workload Input Technique
E - NASA-Task Load Index
F - Post-Trial Participant Subjective Questionnaire
G - Post-Experiment Participant Subjective Questionnaire
H - SME SAGAT Data Evaluation Form

List of Illustrations

Figures Page

Tables Page

Executive Summary

In order to meet industry demands and improve efficiency, the Federal Aviation Administration (FAA) is examining shared-separation responsibility as a future operations concept. Shared-separation responsibility, also referred to as Free Flight, will allow pilots greater freedom to maneuver their aircraft and choose their preferred routes. New technologies and automation could allow pilots to accept greater responsibility for separating their aircraft. Air Traffic Control Specialists (ATCSs) could monitor air traffic and impose restrictions only when necessary to ensure safety. The airline industry may then benefit from more efficient flight operations that will reduce fuel consumption and lower operating costs.

In this shared-separation environment, the role of ATCSs will change from an active command-and-control authority to a more passive element in the system. Allowing pilots greater freedom to maneuver their aircraft may make it more difficult for controllers to maintain situation awareness (SA) and respond to loss of aircraft separation when needed. ATCSs will require new technologies and automation to support their SA and maintain safety in the system. This simulation study evaluated the effects of an advanced display concept on controller SA, workload, and performance while operating in a shared-separation environment. A secondary objective compared different SA measurement techniques and offer recommendations for usage in future air traffic control (ATC) studies.

The advanced display evaluated here consisted of a small window of text information placed on the existing ATC radar display. The display provided information to controllers regarding each aircraft's intended altitude, heading, or airspeed change as pilots freely maneuvered during shared-separation operations. In the study, simulation pilots sent the information when their aircraft initiated a maneuver and removed the entry from the display when their aircraft reached the intended altitude, heading, or airspeed. The advanced display simulated future datalink automation between aircraft flight management systems and ATC facilities.

The shared-separation environment used in the current simulation represented an advanced operations concept with specific rules and assumptions. First, pilots were free to maneuver their aircraft without using voice communications to inform controllers. Controllers were to assume that every aircraft was equipped with an advanced traffic display that allowed pilots to separate their aircraft safely. In addition, controllers were to assume that pilots used an air-to-air voice frequency to communicate with each other and negotiate separation maneuvers. However, the air-to-air frequency was not available for controllers to monitor. Controllers were not to issue traditional ATC commands to pilots. Instead, controllers issued traffic warnings to pilots that identified potential conflicts. Controllers issued advisory separation maneuvers to pilots when necessary to ensure safety, but pilots could ignore these advisories at their own risk.

A team of human factors researchers and ATCSs conducted the simulation in the Research Development and Human Factors Laboratory at the FAA William J. Hughes Technical Center. The study employed the laboratory's high fidelity ATC radar simulation equipment and software. Six simulation pilots communicated with controllers and moved all radar targets using simple keyboard commands. The researchers developed generic en route airspace for the study. The generic airspace represented a high altitude sector that was easy for controllers to learn. The researchers designed traffic scenarios with scripted pilot maneuvers that ensured conflicting

aircraft remained separated. However, each scenario depicted one pair of conflicting aircraft that would not separate safely without controller intervention and an advisory separation maneuver. This situation represented a rare pilot error that illustrated the consequences if controllers did not remain vigilant.

Ten FPL controllers participated in three simulation trials using the current display system and three trials using the advanced display concept. Controller performance was evaluated using objective measures produced by the laboratory's simulation software and with subjective measures provided by the ATCSs using an over-the-shoulder rating form. Controller situation awareness was measured using the Situation Awareness Global Assessment Technique and the Situational Awareness Rating Technique. In addition, the researchers used an on-line SA probe technique that presented questions to controllers during the simulation. Controller workload was assessed using the National Aeronautics and Space Administration Task Load Index and the Air Traffic Workload Input Technique.

The results indicated that the advanced display had some specific benefits for controller performance and SA. Although the results were not statistically significant, controllers tended to make fewer ground-to-air transmissions when using the enhanced display. This result suggests that when controllers received information confirming that pilots were maneuvering correctly for aircraft separation, there was no need to issue further advisories. In addition, the advanced display dramatically increased controller awareness for which aircraft were conforming to their advisories. However, the results indicated that controller workload (specifically mental workload) increased with the advanced display probably due to monitoring additional information.

The results indicated considerable variability between the different SA measurement techniques used in this study. Therefore, the researchers can not make any conclusions regarding the validity of the SA techniques. Ultimately, researchers must use the technique that they believe best represents controller situation awareness. Controllers did indicate that the Situation Awareness Global Assessment Technique may be intrusive to ATC operations, whereas the on-line SA probe technique was not. The researchers recommend additional simulation studies to evaluate other advanced display concepts and their impact on controllers operating in shared-separation environments.

This study represented a relatively low technology implementation of simulated free flight. It demonstrated that a sample of controllers could work with it. As technology and its implementation improve, as projected in Free Flight Phase I, the concept could become increasingly more feasible.

1. Introduction

In order to meet industry demands and improve efficiency, the Federal Aviation Administration (FAA) is examining shared-separation responsibility as a future operations concept. Shared-separation responsibility, also referred to as Free Flight, may allow pilots greater freedom to maneuver their aircraft and choose their preferred routes. New technologies and automation may allow pilots to accept greater responsibility for separating their aircraft. Air Traffic Control Specialists (ATCSs) may monitor air traffic and impose restrictions only when necessary to ensure safety. The airline industry could benefit from more efficient flight operations that will reduce fuel consumption and lower operating costs.

In this shared-separation environment, the role of ATCSs will change from an active command-and-control authority to a more passive element in the system. Allowing pilots greater freedom to maneuver their aircraft may make it more difficult for controllers to maintain situation awareness (SA) and respond to loss of aircraft separation when needed. Controllers will require new technologies and automation to support their situation awareness and maintain safety in the system.

Recent research, however, found that without new types of decision aids, ATCSs may suffer from significant decrements in SA and increased workload (Endsley, Mogford, Allendoerfer, Snyder, & Stein, 1997). These effects may leave controllers unable to adequately back up aircraft that are responsible for self-separation (or to back up automated systems that provide aircraft separation). Controllers may have difficulty detecting and preventing separation problems as their role changes from active controlling to more passive monitoring in a future shared-separation environment.

1.1 Background

The ability of controllers to maintain an up-to-date mental model of a dynamic and complex traffic situation depends on their ability to integrate information about many aircraft into an internal structure. This internalization allows relationships between aircraft to be understood (e.g., which aircraft are traffic for each other, relative speeds, bearings, and ascent or descent rates). This becomes a much more difficult job as the predictability of the aircraft decreases under self-separation conditions. More limited attention and working memory will be required to process each aircraft, leaving the controller with higher workload, poorer SA, or both.

Several factors may be behind these difficulties. The first is a decrease in aircraft movement predictability associated with direct routes. The normal route structure provides all the information necessary for projecting how aircraft will transition across the airspace and with whom they might have future separation problems. This type of projection (the highest level of SA) is extremely important for acting with sufficient timeliness to prevent separation losses. In self-separation ATC environments where aircraft can deviate from their published trajectories at will, ATCSs may be even less able to keep up with air traffic and intervene in a timely manner.

For controllers to maintain an effective role, they need modifications to displays, automated systems, or procedures to regain some degree of predictability. Increasing the predictability should significantly assist them in developing the mental traffic picture needed at a lower level of workload in such operational conditions.

Display enhancements may be more effective than relying on purely automated techniques (e.g., conflict probe) for this assistance. Automated systems can reduce SA through complacency and vigilance effects, greater use of passive processing rather than active processing, and changes in operator feedback (Endsley & Kiris, 1995). Therefore, interface concepts that focus on providing better SA may be more effective for improving operator performance than concepts that focus primarily on decision aiding (Endsley & Selcon, 1997). The goal of the present research was to explore an enhanced display technique designed to provide ATCS with more predictability in processing self-separation air traffic.

1.2 Objectives

The primary objective of this study was to evaluate the effects of an advanced display concept on controller SA, workload, and performance while operating in a shared-separation environment. A secondary objective was to compare different SA measurement techniques and offer recommendations for usage in future ATC studies.

2. Method

2.1 Participants

Participants were all experienced full performance level (FPL) ATCSs from five different en route Air Route Traffic Control Centers (ARTCCs) in the U.S. Ten controllers, eight men and two women, were included in the study. Mean age was 39.6 years. Mean experience was 12.05 years at FPL, with an average of 15.4 years of experience in ATC. All participants were current at their ARTCC with at least 16 hours controlling traffic in the preceding month. All participants were required to have self-reported corrected vision of at least 20/30. Participation in the study was on a voluntary basis and complied with FAA regulations for use of human participants.

2.2 Experimental Design

The study employed a within-participants design with two levels of one independent variable. All participants experienced three trials in each of the two conditions. The researchers provided a counter-balanced order for condition across participants.

2.2.1 Independent Variable

The amount of ATC display support served as the independent variable for the test. We examined two conditions:

- Baseline - This condition employed current ATC radar displays and flight strips. It served as the control condition.

- Enhanced Display - In addition to the traditional radar display and flight strips, this condition incorporated a new display designed to enhance controller SA that appeared in a window on the radar display monitor. The window displayed additional information for each aircraft in the sector that was in a transitionary state (those changing heading, altitude, or airspeed). All displayed aircraft (Table 1) were in a transitionary state (up to the 10 most recent). Each new entry appended to the bottom of the list. As each aircraft reached its target altitude, airspeed, or heading, the displayed entry disappeared. The display provided a simulation of data that might be available in the future air traffic system by datalink from the on-board computers of an aircraft.

Table 1. Enhanced Display

DAL347	R	H120
USA876	L	H090
AWE998	C	A350
UAL641	D	A310

2.2.2 Dependent Variables

The researchers examined several measures as dependent variables, including performance, SA, workload, and the subjective impressions of the controllers serving as participants in this study.

2.2.2.1 Performance

Performance measures included objective measures of controller and system performance and subjective ratings of controller performance.

2.2.2.1.1 Objective Performance Measures

The ATCoach (1992) simulation computer collected a variety of objective performance data during the study. We performed calculations to derive relevant performance measures. They were

a. Safety
   1. Number of En route Conflicts
b. Efficiency
   1. Number of Flights Handled
   2. Number of Flights Completed
   3. Total Aircraft Flight Time
   4. Total Aircraft Distance Flown
c. Communications
   1. Number of Ground-to-Air Transmissions
   2. Duration of Ground-to-Air Transmissions

2.2.2.1.2 Subjective Performance Measures

A subject matter expert (SME) provided a subjective rating of participant performance at the conclusion of each trial using a modified version of the observer checklist developed by Sollenberger, Stein, and Gromelski (1997) (Appendix A). Personnel at the Research Development and Human Factors Laboratory (RDHFL) at the FAA William J. Hughes Technical Center developed the observer checklist used for this study. SMEs for the study were experienced controllers who were familiar with the operational concept and scenarios tested. They observed each participant as they controlled traffic in the scenarios. At the end of each trial, an SME made a subjective rating of the participant performance on each item in the checklist using an 8-point scale. The scale ranged from 1 (extremely poor judgment and made very frequent errors) to 8 (always demonstrated excellent judgment and used outstanding control techniques). The checklist items incorporated certain changes in wording as appropriate for the operational concept. The checklist items were

a. Maintaining Safe and Efficient Traffic Flow
    1. Maintaining Separation and Resolving Potential Conflicts
    2. Using Separation Interventions Effectively
    3. Overall Safe and Efficient Traffic Flow
b. Maintaining Attention and Situation Awareness
    1. Maintaining Awareness of Aircraft Positions
    2. Identifying Traffic Conflict Problems in a Timely Manner
    3. Correcting Own Errors in a Timely Manner
    4. Overall Attention and Situation Awareness
c. Prioritizing
    1. Taking Actions in an Appropriate Order of Importance
    2. Handling Tasks for Several Aircraft
    3. Keeping Data Blocks Up-to-Date
    4. Overall Prioritizing
d. Providing Control Information
    1. Providing Essential Air Traffic Control Information,
    2. Providing Additional Air Traffic Control Information
    3. Overall Providing Control Information
e. Technical Knowledge
    1. Showing Knowledge of Letters of Agreement (LOAs) and Standard Operating Procedures (SOPs)
    2. Showing Knowledge of Aircraft Capabilities and Limitations
    3. Overall Technical Knowledge
f. Communicating
    1. Using Proper Phraseology
    2. Communicating Clearly and Efficiently
    3. Listening for Pilot Readbacks and Requests
    4. Overall Communicating

In addition, the SMEs provided an overall rating describing how well the controller managed traffic on a 10-point scale (1-poor to 10- extremely well) during each trial.

2.2.2.2 Situation Awareness

In addition to the subjective rating of attention and SA provided by the SME on the Observer Checklist, three other measures of SA were included in the study.

2.2.2.2.1 Situation Awareness Global Assessment Technique

The researchers employed the Situation Awareness Global Assessment Technique (SAGAT) as an objective measure of participant SA during the test (Endsley, 1988; Endsley, 1995b; Endsley & Rodgers, 1994). To collect SAGAT data, four freezes occurred in each trial. (Two SAGAT freezes occurred in training trials.) Freezes occurred at randomly selected intervals and were unpredictable to the controller. At the time of the freeze, the controller radar display was blanked and the simulation frozen. A computer collected the controller answers to each of the SAGAT queries at each freeze. SAGAT queries (Appendix B) consisted of questions regarding

a. Level 1 SA - Perception of the Traffic Situation
    1. Aircraft Location
    2. Aircraft Level of Control
    3. Aircraft Callsign
    4. Aircraft Altitude
    5. Aircraft Groundspeed
    6. Aircraft Heading
    7. Aircraft Flight Path Change
    8. Aircraft Type
b. Level 2 & 3 SA - Comprehension and Projection of Traffic Situation
    1. Aircraft Next Sector
    2. Aircraft Separation
    3. Aircraft Advisories
    4. Advisory Reception
    5. Advisory Conformance
    6. Aircraft Hand-offs
    7. Aircraft Communications
    8. Special Airspace Separation
    9. Weather Impact

2.2.2.2.2 On-line SA Probes

The researchers also measured SA using an on-line probe technique that presented questions to controllers during the simulation. The on-line SA probe technique is similar to the Situation Present Assessment Method (SPAM) developed by Durso et al. (1998). This technique allowed SA data to be collected without requiring freezes in the simulation.

Throughout the trial, one of the experimenters posing as the controller of an adjacent sector periodically interjected SA probes while the simulation was running. These probes took the form of verbal questions that directly corresponded to eight of the SAGAT queries (four Level 1 SA probes and four Level 2/3 SA probes). The on-line SA probes were similar in nature to the SAGAT queries, allowing a comparison between measures. However, they occurred while the simulation was running and the participant had full view of the displays and flight strips. The researchers set the timing of these SA probes either 1 minute before or 1 minute after each workload rating probe, beginning at a point ten minutes into the trial. This timing allows a comparison of the on-line SA probe reaction times (RT) with the corresponding workload probe. The researchers administered the SA probes in a random order and recorded the accuracy of the participant response and time to respond. The following list identifies the probes.

a. Level 1 SA - Perception of Traffic Situation
    1. What is the current heading for aircraft X?
    2. What is the current flight level for aircraft X?
    3. Climbing, descending or level: which is correct for aircraft X?
    4. Turning right, turning left, or on course: which is correct for aircraft X?
b. Level 2 & 3 SA - Comprehension and Projection of Traffic Situation
    1. Which aircraft have lost or will lose separation if they stay on their current (intended) course?
    2. Which aircraft will be affected by weather within the next 5 minutes unless an action is taken to avoid it?
    3. Which aircraft must be handed off within the next 3 minutes?
    4. What is the next sector for aircraft X?

2.2.2.2.3 Situational Awareness Rating Technique

Finally, SA was also measured using the Situational Awareness Rating Technique (SART) (Taylor, 1990). The 10-dimension version of SART provides a bipolar scale for participants to subjectively rate their own SA on each of 10 items (shown in Appendix C). The 10 ratings combine to form a rating for each of three major factors: supply of resources, demand for resources, and understanding. An overall SART rating forms from these factors. The researchers obtained SART at the completion of each trial with a computer.

2.2.2.3 Workload

The participants provided three measures of workload, and the SMEs provided two subjective measures of workload. The researchers administered the Air Traffic Workload Input Technique (ATWIT) at 5-minute intervals throughout the trial to obtain a subjective workload rating from the participants on a 10-point scale (Appendix D). ATWIT provides an unobtrusive and reliable means for collecting controllers' workload ratings (Stein, 1985; Stein, 1991). The participants and the SMEs also provided a subjective workload rating immediately following each trial using the National Aeronautics and Space Administration (NASA) Task Load Index (NASA-TLX) (Hart & Staveland, 1988), provided in Appendix E. Finally, the SMEs and the participants provided a subjective assessment of workload on a 10-point scale.

2.2.2.4 Subjective Questionnaire

The researchers provided a Post-Trial Participant Subjective Questionnaire (Appendix F) to each participant after each trial. On this questionnaire, the participant evaluated how hard they were working, how difficult the scenario was, and how well they controlled traffic in the preceding trial. The researchers distributed a Post-Experiment Participant Subjective Questionnaire (Appendix G) at the end of the study. Participants provided their opinion on the study, the enhanced display, and the level of perceived intrusiveness and ease of use of the SA and workload measures included in the study.

2.3 Apparatus

The researchers conducted the test in the RDHFL, which is a high-fidelity ATC simulation facility with equipment that includes controller workstations with high-resolution radar displays, three-button trackballs, ATC keyboards, and ATCoach (1992) simulation software. A voice communication link allowed controllers to issue commands to a remote room of simulation pilots who moved the radar targets. A flight strip bay with printed, standard-configuration flight strips provided information for each aircraft in the simulation. A panel for the ATWIT measure was included to the left of the radar display. The researchers administered SAGAT, SART and NASA-TLX procedures using a computer placed adjacent to the controller's station. The researchers provided on-line SA probes verbally over the participant headsets and subjective measures using paper forms.

2.4 Test Scenarios

The researchers created 10 scenarios for the simulation. These included six test scenarios and four training scenarios of approximately 45 minutes each. The six scenarios were similar to each other in terms of relative traffic density and complexity. The scenarios took place in a generic airspace high altitude sector (Genera Sector), which was developed and validated for conducting human factors research testing (Gutman, Stein, & Gromelski, 1995).

The scenarios used a self-separation operational concept. Under these flight rules, the pilot and the ATCS each had certain responsibilities and assumptions.

a. The pilot
    1. was responsible for aircraft safety and in-flight separation with other aircraft;
    2. was responsible for changing routes and altitudes to ensure aircraft separation;
    3. did not need to inform ATC of actions beforehand but only communicate with ATC as needed for coordination (e.g., respond to queries);
    4. assumed having onboard systems to provide weather, special use airspace, and other aircraft information;
    5. may communicate with other pilots as needed (through voice or datalink);
    6. assumed that a datalink-like system would downlink certain aircraft information to ATC and other aircraft (e.g., that incorporated in the ATC enhanced display concept).
    7. may choose to follow the advisory or elect to take another course if desired to resolve the conflict (advisories and warnings provide advice and not a commanded clearance as in the current system);
    8. retains responsibility for the aircraft separation at all times. The pilots can choose another course of action (override the controller suggestion) at their own peril but may choose to do so as they are always responsible for the safety and flight path of the aircraft.
b. The ATCS
    1. was responsible for normal sector hand-off/point-out procedures and coordination;
    2. must have the data block up to date before aircraft was handed off to next sector;
    3. provided back up for aircraft separation. This means
        a) if the controller determines that two aircraft are in conflict and have not self-separated, the controller is to step in and resolve the conflict by
            1) giving warning - (e.g., traffic at 2 o’clock), or
            2) providing an advisory - (e.g., advise descend and maintain 10,000 feet).

b) the controller should provide the warning, followed by the advisory if needed. The controller should provide the warning or advisory to both aircraft involved.

c) if the controller sees that an aircraft actions (deviations from their listed flight plan) will lead them into conflict with a third aircraft, they should provide a warning or advisory as needed to that aircraft.

According to the type of performance expected under self-separation (assuming some form of on-board aiding system), the scenarios were scripted so that the aircraft maintained the required separation or better between aircraft during the scenarios. During each scenario, however, a few problems were included in which the aircraft did not adequately maintain separation. In these cases, controller intervention was required to insure separation. Controllers did not know in which cases the pilots would separate themselves and in which cases they would not do so adequately. The scenarios, therefore, represented an operational concept in which controllers occupy a passive monitoring role, intervening primarily by exception.

The scenarios incorporated hand offs, procedures, and separation standards as are currently employed in controlling traffic with the exception of the procedural changes described. Flight plans for the aircraft used direct routing. Poor weather and turbulence conditions in the scenarios provided rationale for pilot variance from direct routes and realism in the simulation.

2.5 Procedures

Each participant controlled traffic in both conditions, administered in a counter-balanced order. The three trials for each condition occurred consecutively. Evenly assigned to each condition, the six scenarios occurred in a counter-balanced order across trials and participants.

Each participant completed the study over 3 consecutive days. A lunch break and two rest breaks occurred each day. On the initial day of the study, participants received an introduction to the simulator and study instructions. They also received instructions for SAGAT, the on-line SA probes, SART, ATWIT, and NASA-TLX. The participants completed the required NASA-TLX paired-comparison rating forms. They received 1 hour of familiarization training on the ATCoach (1992) simulator, Genera Sector, and the operational concept.

For each condition, participants received two training trials on the experimental condition tested. During each of two stops to practice filling out the SAGAT, there were four on-line SA probes. Three trials for that condition followed. An audio tone at five-minute intervals prompted participants to fill out the ATWIT scale. The eight on-line SA probes occurred over the participant headsets either 1 minute before or 1 minute following each ATWIT rating. The researchers instructed the participants to respond verbally to each on-line SA probe as quickly and accurately as possible.

Four freezes occurred in each scenario at random times to collect SAGAT data. At the time of each freeze, the radar screen blanked and the simulation paused while the participant completed the SAGAT queries on the computer. The participants first saw an electronic map of the sector that showed only the boundaries and navigation fix points. The researchers asked them to indicate the location on the map of all aircraft currently under their control, all aircraft handed-off but still inside their sector boundaries, and all aircraft soon to be under their control. The remaining SAGAT queries then appeared in a random order in relation to the aircraft that the participants indicated were present on the map. Participants completed all queries and then returned to the simulation at the point where they had left off (taking approximately 5 minutes for the freeze). The researchers gave them a few seconds to observe the radar screen, after which they resumed the simulation.

As the participants filled out the SAGAT battery, the SME filled out the SAGAT data collection form (Appendix H) while viewing the radar data and flight strips. This supplemented the data collected by the simulation computer. At the end of each trial, the SME filled out the observation checklist form rating the participant performance. The participants completed the SART, NASA-TLX, and Post-Scenario Questionnaire at the end of each trial. They also completed the Post-Experiment Questionnaire at the end of all of the simulation trials.

3. Results

The results of this study are in four main sections: ATCS performance, controller workload, controller SA, and subjective evaluations. In addition, the researchers conducted a comparative analysis of the SA measures used in the study to assess the utility of these tools for ATC research.

Analysis of Variance (ANOVA) including main effects of condition, trial, and the two-way interaction. The Tukey-HSD procedure post hoc test provided comparisons between conditions. The researchers used an alpha level of $p < .05$ for all statistical tests.

3.1 Air Traffic Control Performance Results

Table 2 shows the ANOVA results for the objective measures of performance, and Table 3 shows the ANOVA results for the subjective measures of performance. The researchers discuss these results separately.

Table 2. Analysis of Variance Results for Objective Performance Measures

	Condition		**Trial**		**Interaction**	
Variable	***F*-Value**	**df**	***F*-Value**	**df**	***F*-Value**	**df**
Safety						
Number of Standard En route Conflicts	0.130, N.S.	1,9	0.550, N.S.	2,18	2.670, N.S.	2,18
Efficiency						
Number of Flights Handled	0.000, N.S.	1,9	0.730, N.S.	2,18	4.850*	2,18
Number of Flights Completed	8.620*	1,9	0.100, N.S.	2,18	0.340, N.S.	2,18
Total Aircraft Flight Time	0.080, N.S.	1,9	1.250, N.S.	2,18	0.910, N.S.	2,18
Total Aircraft Distance Flown	0.060, N.S.	1,9	1.050, N.S.	2,18	1.010, N.S.	2,18
Communications						
Number of Ground-to-Air Transmissions	3.990, N.S.	1,9	2.060, N.S.	2,18	1.390, N.S.	2,18
Duration of Ground-to-Air Transmissions	3.680, N.S.	1,9	3.130, N.S.	2,18	0.850, N.S.	2,18

*$p < .05$

Table 3. *F* Statistics Obtained from the Two-way ANOVA Performed on the Observer and Participant Performance Ratings

Measure	Main Effect: Condition	Main Effect: Trial	Interaction Effect
Overall Performance Ratings			
Participant Overall Rating	F (1, 9) = 0.39	F (2, 18) = 0.75	F (2, 18) = 2.45
Observer Overall Rating	F (1, 9) = 0.93	F (2, 18) = 0.73	F (2, 18) = 15.13**
Observer Ratings			
Safe & Efficient Traffic Flow			
Maintaining separation	F (1, 9) = 0.22	F (2, 18) = 0.77	F (2, 18) = 4.51*
Using control instructions	F (1, 9) = 0.87	F (2, 18) = 0.48	F (2, 18) = 3.25
Overall	F (1, 9) = 0.79	F (2, 18) = 0.38	F (2, 18) = 3.75*
Attention & Situation Awareness			
Maintaining awareness	F (1, 9) = 2.09	F (2, 18) = 0.47	F (2, 18) = 5.46*
Determining traffic problems	F (1, 9) = 1.71	F (2, 18) = 1.78	F (2, 18) = 5.16*
Correcting own errors	F (1, 9) = 5.85*	F (2, 18) = 2.06	F (2, 18) = 6.05**
Overall	F (1, 9) = 2.12	F (2, 18) = 0.39	F (2, 18) = 8.48**
Prioritizing			
Taking actions in order	F (1, 9) = 0.21	F (2, 18) = 0.63	F (2, 18) = 5.95*
Handling tasks for several aircraft	F (1, 9) = 0.17	F (2, 18) = 2.65	F (2, 18) = 3.09
Keeping datablocks up-to-date	F (1, 9) = 0.10	F (2, 18) = 0.26	F (2, 18) = 2.98
Overall	F (1, 9) = 0.01	F (2, 18) = 0.68	F (2, 18) = 3.31
Providing Control Information			
Providing essential info	F (1, 9) = 1.01	F (2, 18) = 1.98	F (2, 18) = 6.38**
Providing additional info	F (1, 9) = 0.02	F (2, 18) = 5.25*	F (2, 18) = 13.05**
Overall	F (1, 9) = 1.51	F (2, 18) = 1.07	F (2, 18) = 7.28**
Technical Knowledge			
Knowing LOAs and SOPs	F (1, 9) = 2.86	F (2, 18) = 1.42	F (2, 18) = 2.36
Knowing aircraft capabilities	F (1, 9) = 2.67	F (2, 18) = 3.00	F (2, 18) = 1.87
Overall	F (1, 9) = 1.71	F (2, 18) = 1.98	F (2, 18) = 4.77*
Communicating			
Using proper phraseology	F (1, 9) = 1.02	F (2, 18) = 4.29*	F (2, 18) = 2.76
Communicating clearly	F (1, 9) = 1.98	F (2, 18) = 0.45	F (2, 18) = 6.23**
Listening to pilots	F (1, 9) = 0.76	F (2, 18) = 0.35	F (2, 18) = 2.94
Overall	F (1, 9) = 0.76	F (2, 18) = 1.09	F (2, 18) = 4.28*

* indicates a statistically reliable effect at a significance level of $p < .05$
** indicates a statistically reliable effect at a significance level of $p < .01$

3.1.1 Objective Performance Measures

No significant main effects or interaction effects were present for any of the safety of flight measures shown in Table 2. Of the efficiency measures, the display condition significantly effected the number of completed flights. More flights were completed in the enhanced display condition (mean = 23.37 flights) as compared to the baseline condition (mean = 22.37 flights). The number or duration of flights handled, the distance the aircraft traveled in the sector, and the cumulative average aircraft density were not significantly different between conditions. Further examination showed that there was a significant condition by trial interaction effecting the number of flights handled. While the number of aircraft handled by the third trial was slightly higher in the baseline condition, it was slightly lower in the enhanced display condition, as shown in Figure 1.

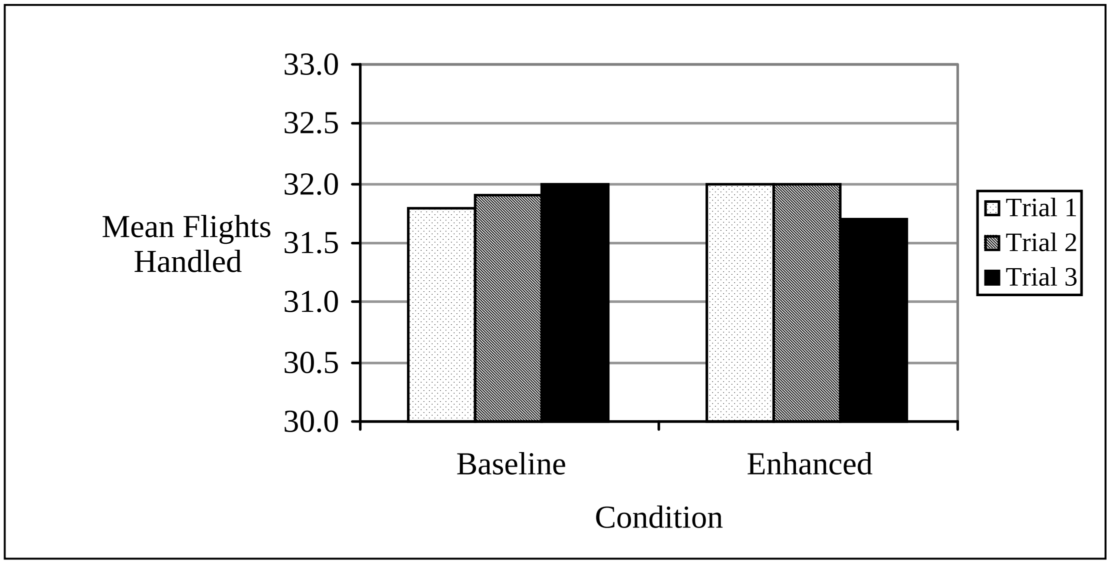

Figure 1. Mean flights handled by condition and trial.

None of the control strategy or communication and taskload variables showed significant main effects or interaction effects. No use of speed as a control strategy occurred during the study. There was a trend (p = .057) towards fewer heading changes per aircraft handled in the enhanced condition (.20) as compared to the baseline condition (.57). Controllers also tended towards fewer transmissions (105.7 vs. 126.0) of shorter duration (364.3 vs. 457.1) in the enhanced condition as compared to the baseline condition, although these were not significant at the $p < .05$ level.

3.1.2 Subjective Performance Measures

Controllers provided a subjective evaluation of their own performance at the end of each trial. Subjective ratings of how well they controlled traffic were not significantly different between conditions or trials as shown in Table 3. No significant difference between conditions or trials in the SME ratings of overall controller performance was present, although there was a significant condition by trial interaction. An examination of the interaction revealed that it reflects differences between the two SMEs who served as raters in the study. As the SMEs alternated serving as the rater for each controller participant after each trial, the interaction reflected the fact that one of the SMEs tended to provide subjective ratings that were slightly higher than the other SME. Similarly, a significant condition by trial interaction existed for many of the subjective ratings shown in Table 3. Because each SME rated each participant an equal number of times, this effect should not have confounded the other results. As the interaction reflects differences in the raters rather than an issue of interest to the study, the interactions in the subjective performance ratings will not be further discussed.

There were no significant main effects for the SME ratings related to safe and efficient traffic flow, prioritizing, or technical knowledge shown in Table 3. Analysis of the SME ratings relating to maintaining attention and SA revealed that ratings of the degree to which controllers corrected their own errors in a timely manner were better in the enhanced display condition, as shown in Figure 2. Trial or condition did not significantly affect other ratings in this category.

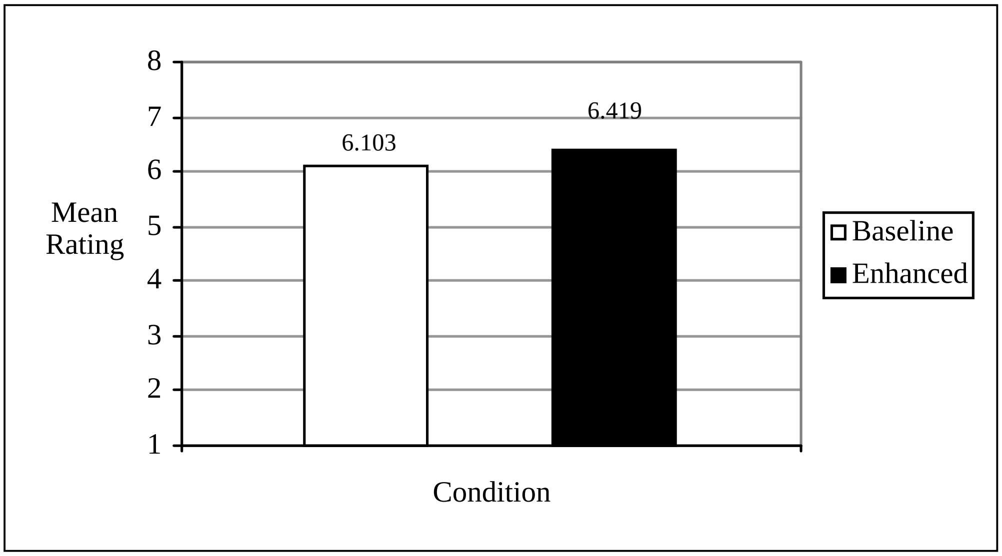

Figure 2. SME ratings of participants correcting own errors in a timely manner.

Analysis of the SME ratings of controller performance revealed no significant difference between conditions in the ratings associated with providing control information or communicating. Two variables showed a significant main effect of trial, however. Trial significantly affected providing additional ATC information, as shown in Figure 3. As a significant trial by condition interaction also was present on this variable, the researchers evaluated the effect of trial for each of the two conditions separately. The effect of trial was significant for both the baseline display condition [$F(2, 17) = 12.358$, $p < .001$], and the enhanced display condition [$F(2, 19) = 6.819$, $p = .006$]. Figure 3 shows that ratings were the highest in the second trial for the baseline condition but were the highest in the third trial for the enhanced condition. The difference between the two raters primarily appears to affect this pattern. Trial was also a significant main effect for ratings on using proper phraseology as shown in Figure 4. Ratings were slightly higher in the third trial as compared to the first two trials.

3.2 Controller Workload

Each participant subjectively rated workload using NASA-TLX, ATWIT, and a post-scenario 10-point scale. The SMEs also subjectively rated controller workload using NASA-TLX and a post-scenario 10-point scale. Results of the ANOVA for condition, trial, and the condition-by-trial interaction effect on the workload measures are in Table 4.

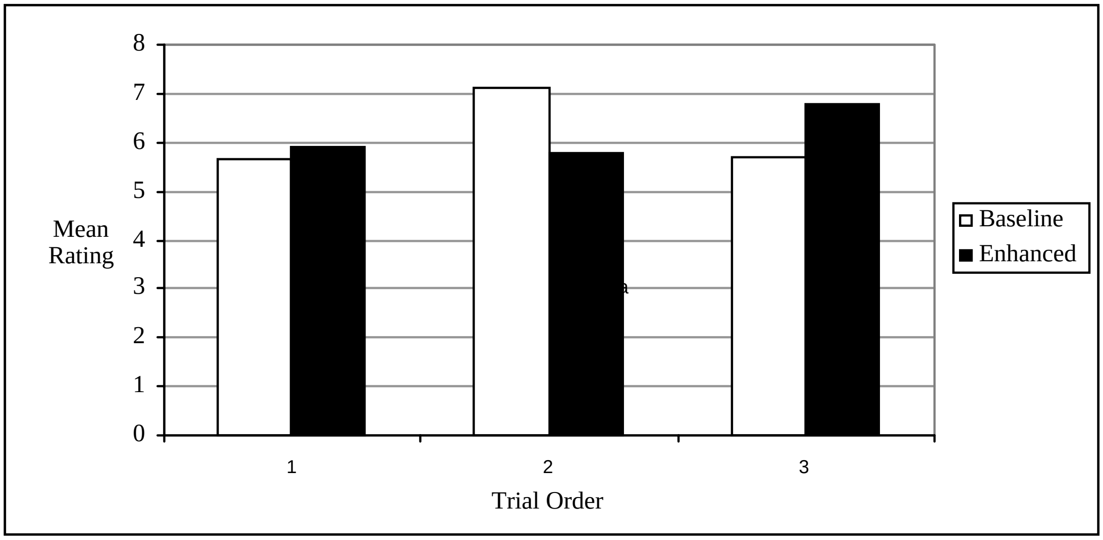

Figure 3. Providing additional air traffic control information by trial and condition.

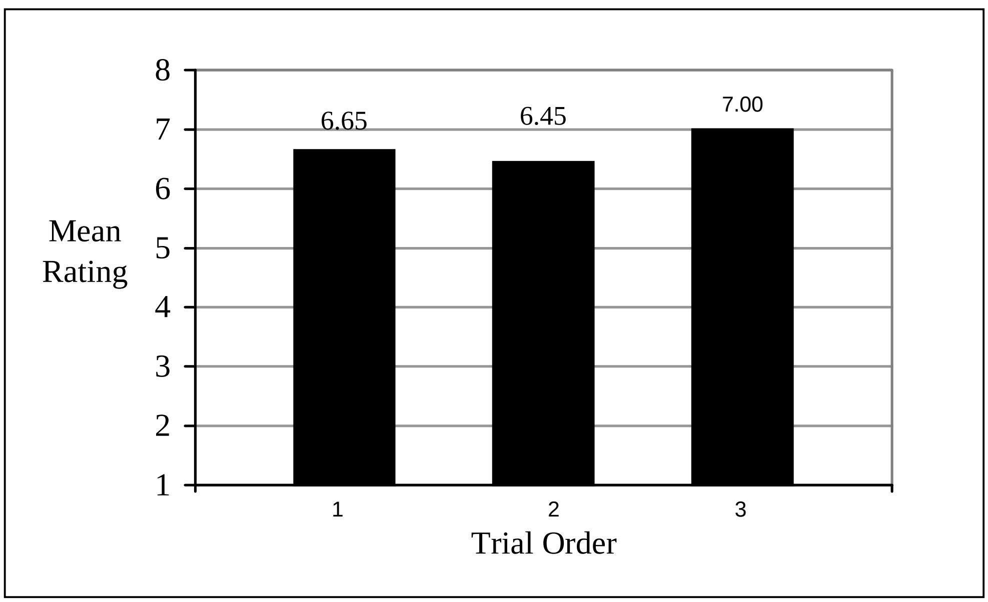

Figure 4. Using proper phraseology by trial.

Table 4. *F* Statistics Obtained from the Two-way ANOVA Performed on the Workload Measures

Measure	Main Effect: Condition	Main Effect: Trial	Interaction Effect
Participant Ratings			
How hard I was working	F (1, 9) = 0.52	F (2, 18) = 1.17	F (2, 18) = 0.21
NASA-TLX Overall Rating	F (1, 9) = 0.33	F (2, 18) = 1.30	F (2, 18) = 0.34
NASA-TLX Mental Load	F (1, 9) = 23.10**	F (2, 18) = 1.90	F (2, 18) = 0.19
NASA-TLX Physical Load	F (1, 9) = 2.10	F (2, 18) = 3.43	F (2, 18) = 1.17
NASA-TLX Temporal	F (1, 9) = 0.64	F (2, 18) = 0.69	F (2, 18) = 0.61
NASA-TLX Performance	F (1, 9) = 0.01	F (2, 18) = 0.09	F (2, 18) = 0.19
NASA-TLX Effort	F (1, 9) = 0.74	F (2, 18) = 0.68	F (2, 18) = 0.13
NASA-TLX Frustration	F (1, 9) = 0.72	F (2, 18) = 0.96	F (2, 18) = 2.00
Mean ATWIT Rating	F (1, 9) = 0.17	F (2, 18) = 1.83	F (2, 18) = 2.11
Observer Ratings			
How hard controller was working	F (1, 9) = 0.75	F (2, 18) = 1.05	F (2, 18) = 2.06
NASA-TLX Mental Load	F (1, 9) = 0.16	F (2, 18) = 0.01	F (2, 18) = 0.51
NASA-TLX Physical Load	F (1, 9) = 0.90	F (2, 18) = 0.50	F (2, 18) = 3.97*
NASA-TLX Temporal	F (1, 9) = 0.71	F (2, 18) = 0.16	F (2, 18) = 0.66
NASA-TLX Performance	F (1, 9) = 1.77	F (2, 18) = 0.58	F (2, 18) = 11.37**
NASA-TLX Effort	F (1, 9) = 0.00	F (2, 18) = 0.10	F (2, 18) = 0.96
NASA-TLX Frustration	F (1, 9) = 0.80	F (2, 18) = 1.42	F (2, 18) = 3.43

* indicates a statistically reliable effect at a significance level of $p < .05$
** indicates a statistically reliable effect at a significance level of $p < .01$

3.2.1 Controller Self-Ratings of Workload

Analysis of the ATWIT mean and peak ratings and the time to respond to the ATWIT tone revealed no significant main effects or interaction effect. Figure 5 depicts a graph of the mean ATWIT ratings across the 5-minute intervals in the trial for each of the display conditions. Whereas there were some variations in ATWIT ratings, with lower ratings at the beginning and end of the trial, overall differences between the conditions were not significant. End-of-the-trial controller ratings of how hard they were working and the difficulty of the trial were not significantly different between conditions, trials, or the trial-by-condition interaction.

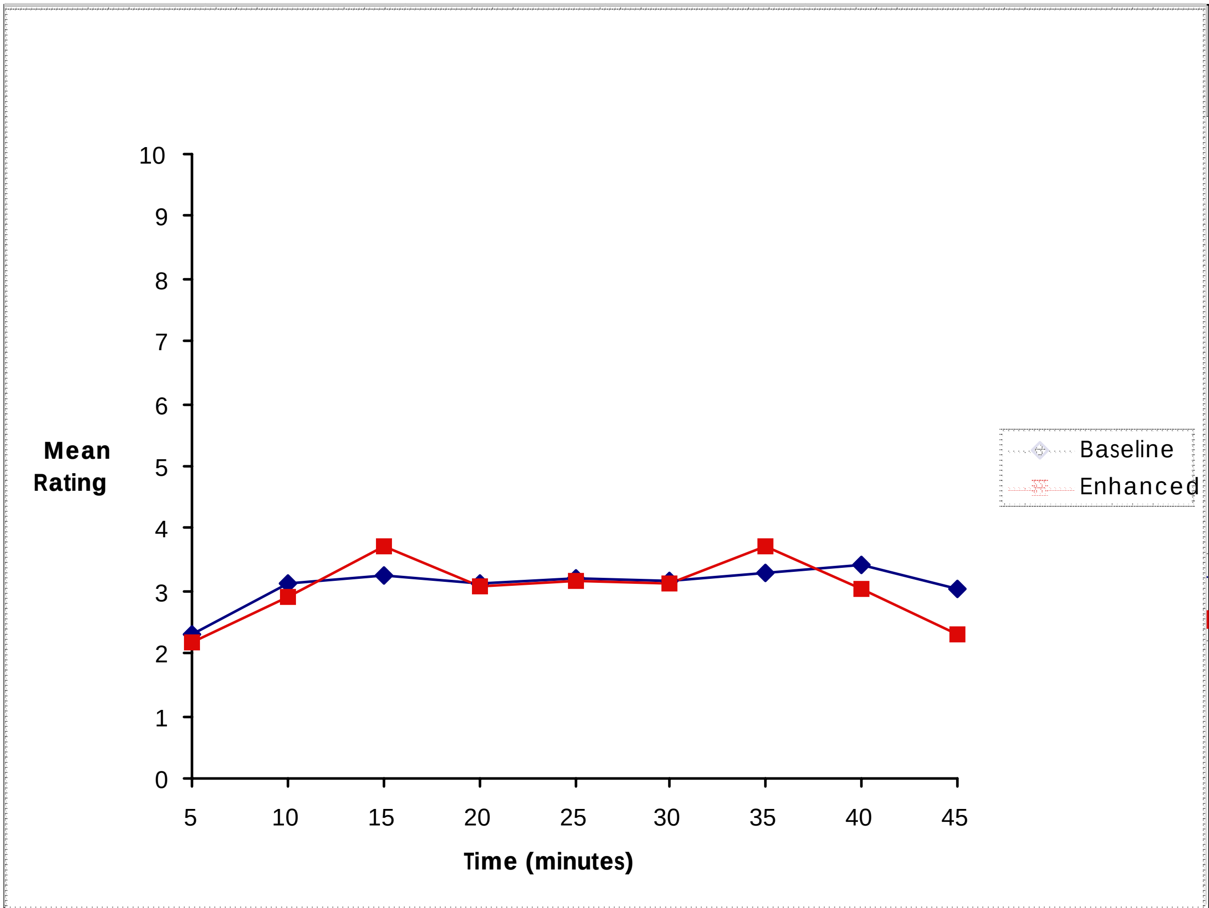

Figure 5. Mean ATWIT ratings across trial.

The overall NASA-TLX rating was also not significantly different in main effects or the interaction. However, closer analysis revealed a significant display condition effect in the mental workload component. Mental workload was rated as higher under the enhanced display condition (mean = 4.9) than in the baseline condition (mean = 4.2).

3.2.2 Subject Matter Expert Ratings of Controller Workload

Analysis of SME ratings of controller workload on the 10-point scale and on the NASA-TLX scales revealed no significant differences between conditions or trial but a significant trial by condition interaction for the physical load and performance subscales. An examination of the data revealed that these differences primarily reflected the difference between the two raters.

3.3 Controller Situation Awareness

Controller SA was measured objectively using SAGAT and the on-line SA probes and subjectively using SART and the SME ratings. Table 5 shows the results of the ANOVAs for the SA measures.

Table 5. *F* Statistics Obtained from the Two-way ANOVA Performed on the Situation Awareness Measures

Measure	Main Effect: Condition	Main Effect: Trial	Interaction Effect
SAGAT			
Aircraft in Sector Location	F (1, 9) = 0.01	F (2, 18) = 1.15	F (2, 18) = 0.24
All Aircraft Location	F (1, 9) = 0.05	F (2, 18) = 0.69	F (2, 18) = 0.09
Aircraft Level of Control	F (1, 9) = 1.08	F (2, 18) = 1.13	F (2, 18) = 0.50
Callsign – Alphabetic Portion	F (1, 9) = 0.39	F (2, 18) = 0.22	F (2, 18) = 0.30
Callsign – Numeric Portion	F (1, 9) = 0.00	F (2, 18) = 1.47	F (2, 18) = 0.80
Altitude	F (1, 9) = 0.04	F (2, 18) = 0.83	F (2, 18) = 4.20*
Speed	F (1, 9) = 1.55	F (2, 18) = 1.69	F (2, 18) = 0.02
Heading	F (1, 9) = 0.51	F (2, 18) = 0.36	F (2, 18) = 0.06
Vertical Change	F (1, 9) = 4.09	F (2, 18) = 0.39	F (2, 18) = 0.06
Turning	F (1, 9) = 1.14	F (2, 18) = 1.91	F (2, 18) = 0.20
Aircraft Type	F (1, 9) = 0.50	F (2, 18) = 0.57	F (2, 18) = 0.37
Next Sector	F (1, 9) = 1.19	F (2, 18) = 0.57	F (2, 18) = 1.17
Aircraft Separation	F (1, 9) = 0.07	F (2, 18) = 1.12	F (2, 18) = 0.43
Advisory Reception	F (1, 9) = 4.21	F (2, 18) = 0.92	F (2, 18) = 0.16
Advisory Conformance	F (1, 9) = 6.96*	F (2, 18) = 0.34	F (2, 18) = 0.08
Aircraft Handoffs	F (1, 9) = 0.00	F (2, 18) = 2.80	F (2, 18) = 0.47
Aircraft in Communication	F (1, 9) = 0.58	F (2, 18) = 0.22	F (2, 18) = 0.50
Special Airspace Separation	F (1, 9) = 1.43	F (2, 18) = 0.03	F (2, 18) = 0.82
Weather Impact	F (1, 9) = 0.39	F (2, 18) = 0.46	F (2, 18) = 1.98
On-line SA Probes			
Vertical Change Probe RT	F (1, 9) = 0.56	F (2, 18) = 0.40	F (2, 18) = 0.03
Altitude Probe RT	F (1, 9) = 2.66	F (2, 18) = 0.73	F (2, 18) = 0.72
Heading Probe RT	F (1, 9) = 2.20	F (2, 18) = 0.20	F (2, 18) = 2.04
Turning Probe RT	F (1, 9) = 0.13	F (2, 18) = 1.82	F (2, 18) = 1.06
Hand-off Probe RT	F (1, 9) = 0.16	F (2, 18) = 0.98	F (2, 18) = 0.15
Next Sector Probe RT	F (1, 9) = 0.88	F (2, 18) = 0.15	F (2, 18) = 0.72
Separation Probe RT	F (1, 9) = 0.72	F (2, 18) = 0.27	F (2, 18) = 0.69
Weather Probe RT	F (1, 9) = 1.33	F (2, 18) = 0.20	F (2, 18) = 0.43
SART			
Overall SART Rating	F (1, 9) = 1.57	F (2, 18) = 2.00	F (2, 18) = 5.67*
Demand Rating	F (1, 9) = 0.16	F (2, 18) = 0.32	F (2, 18) = 2.77
Supply Rating	F (1, 9) = 0.91	F (2, 18) = 1.64	F (2, 18) = 0.66
Understanding Rating	F (1, 9) = 1.50	F (2, 18) = 2.10	F (2, 18) = 2.55

* indicates a statistically reliable effect at a significance level of $p < .05$
** indicates a statistically reliable effect at a significance level of $p < .01$

3.3.1 Situation Awareness Global Assessment Technique

The participant perception of the traffic situation as reported on the SAGAT queries were compared to the actual state of the traffic situation at the time of each freeze. The researchers scored their answers as correct or incorrect and used an arcsine transformation to correct for non-normality of binomial data. The ANOVAs revealed no significant main effects of trial or condition on any of the Level 1 SA queries nor on a combined measure composed of the Level 1 SA queries. A significant trial by condition interaction was present, however, for controller knowledge of aircraft altitude (Figure 6). The interaction reflects that in the baseline condition, the controller scores were higher in the second trial but, in the enhanced condition, it was lower in the second trial.

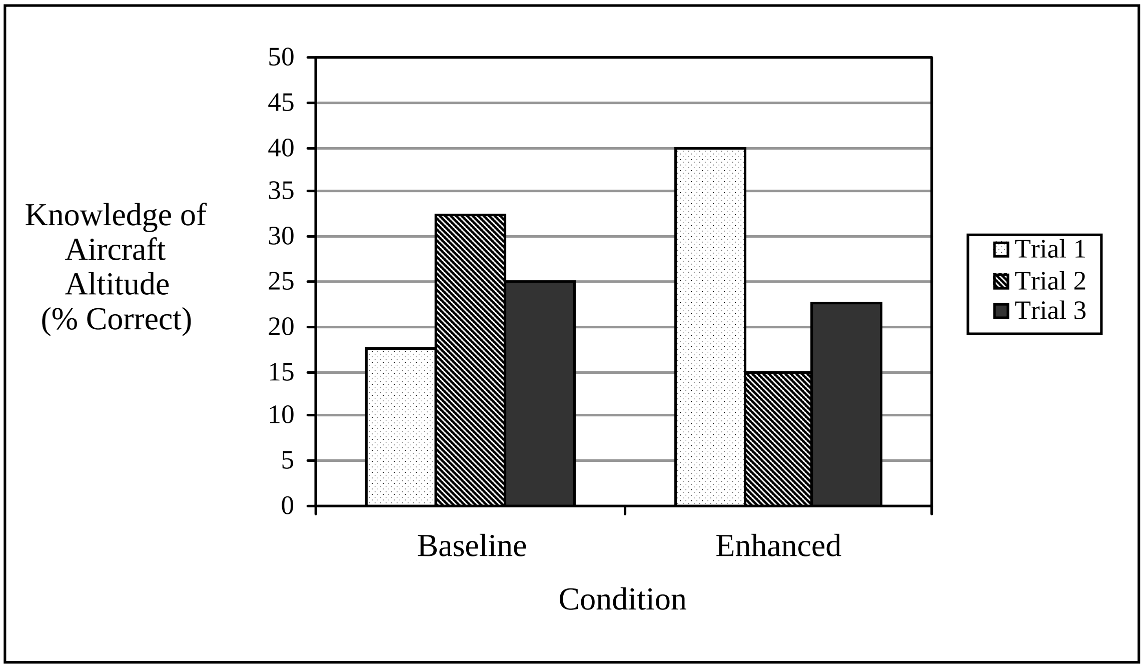

Figure 6. Condition by trial interaction on knowledge of aircraft altitude.

Of the Level 2 SA queries, the researchers found a significant difference between conditions in the participant knowledge of aircraft conformance to advisories. As shown in Figure 7, controllers were over three times more likely to be correct in understanding whether aircraft were conforming to their advisories with the enhanced display. The researchers found no other differences between conditions or trials on the Level 2/3 SA SAGAT queries or on a combined measure composed of the Level 2/3 queries.

3.3.2 On-line Situation Awareness Probes

Accuracy on the on-line SA probes was very high (95% overall), as would be expected. The accuracy was not significantly effected by question [$F(7, 63) = 1.754$, $p = .113$] or condition [$F(1, 9) = .201$, $p = .662$]. The researchers analyzed the RT to the correct ON-LINE probes to determine whether scores were effected by condition [$F(1, 455) = 2.807$, $p = .095$], probe type [$F(7, 455) = 10.422$, $p < .001$], or the condition-by-probe type interaction [$F(7, 455) = 1.237$, $p = .281$]. Figure 8 shows that SA probe type significantly effected probe RT. The SA probe regarding aircraft separation appeared to take the longest on average to answer, and the probe concerning the vertical profile of the aircraft (climbing, descending, or level) took the least amount of time.

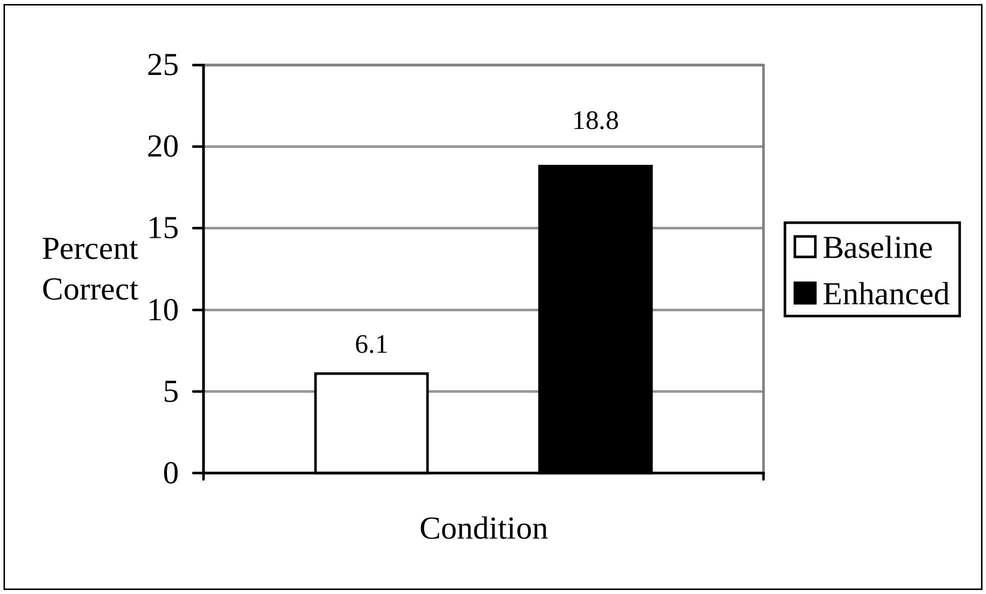

Figure 7. Knowledge of aircraft conformance to advisories by condition.

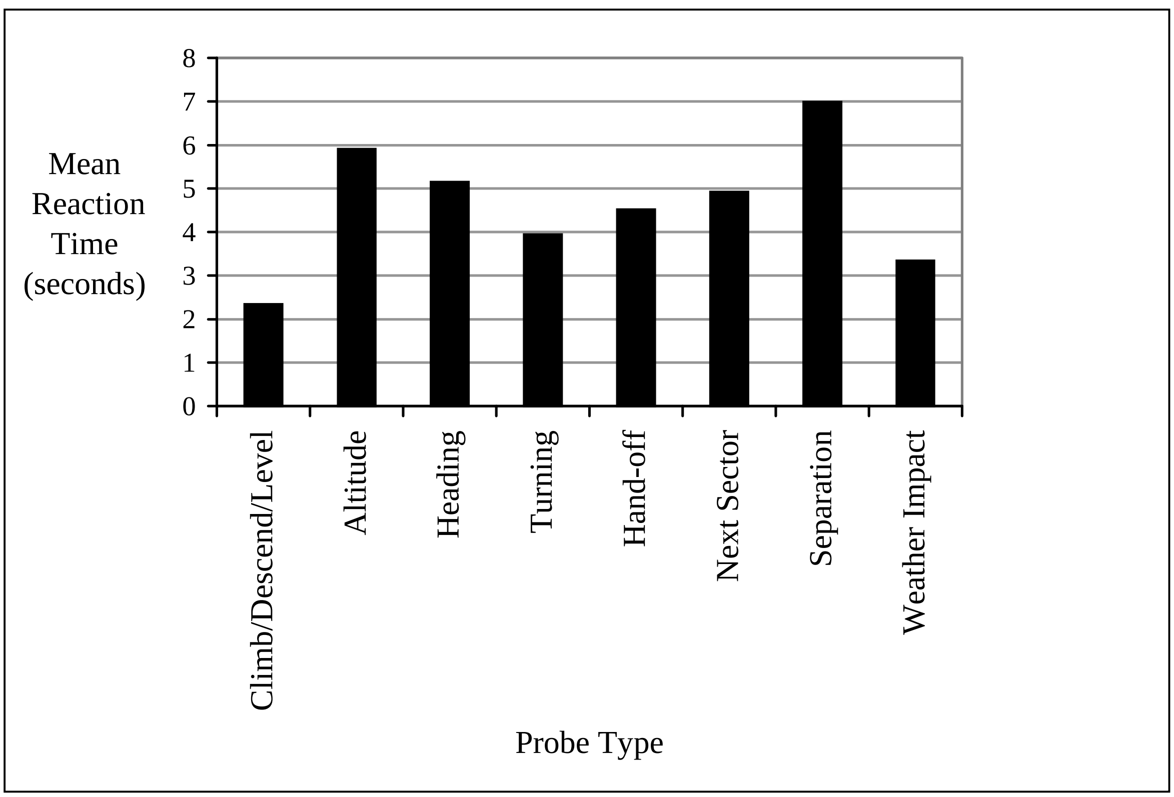

Figure 8. Mean SA probe reaction time by probe type.

Because the SA probe type was a significant factor, the researchers analyzed RT to each probe type separately to examine effects due to condition, trial, and the condition-by-trial interaction. These results, shown in Table 5, revealed no significant main effects or interaction effects. There were also no significant main effects or interaction effect for the Level 1 SA probes combined (present situation) or for the Level 2/3 SA probes combined (future situation).

3.3.3 Situational Awareness Rating Technique

The subjective ratings of the participants on the 10 SART scales combined to form a rating for supply of resources, demand for resources, and understanding. The researchers calculated an overall SART rating from the mean rating of understanding, plus the supply of resources, minus the demand for resources. No significant main effects of trial or condition resulted for the ANOVA of overall SART rating or its three subscales. Figure 9 shows a significant trial by condition interaction occurred for the overall SART rating. An examination of the data showed that the mean SART rating was higher on the first trial with the enhanced display. However, this effect diminished in subsequent trials.

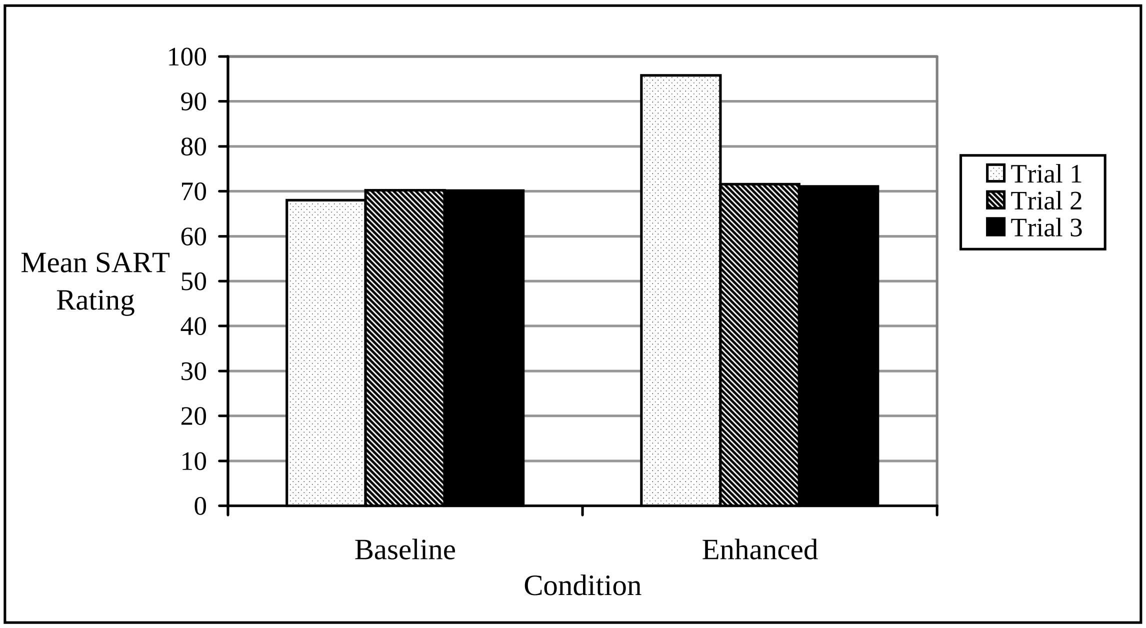

Figure 9. Mean SART rating by condition and trial.

3.3.4 Subject Matter Expert Rating of Situation Awareness and Attention

The SMEs rated four factors related to controller SA and attention including an overall rating. None of these factors was significantly different across conditions based on the ANOVAs.

3.4 Controller Subjective Evaluations

Controllers provided a subjective evaluation of the conditions experienced during the study following their participation (Figure 10). Overall, they rated the realism of the scenarios as only 4.6 on a 10-point scale, reflecting their perceptions of the artificial nature of the self-separation concept. They rated training and the responsiveness of the simulation pilots higher, however. The enhanced display received a mean rating of 5.67 on a 10-point scale, showing a moderate level of utility.

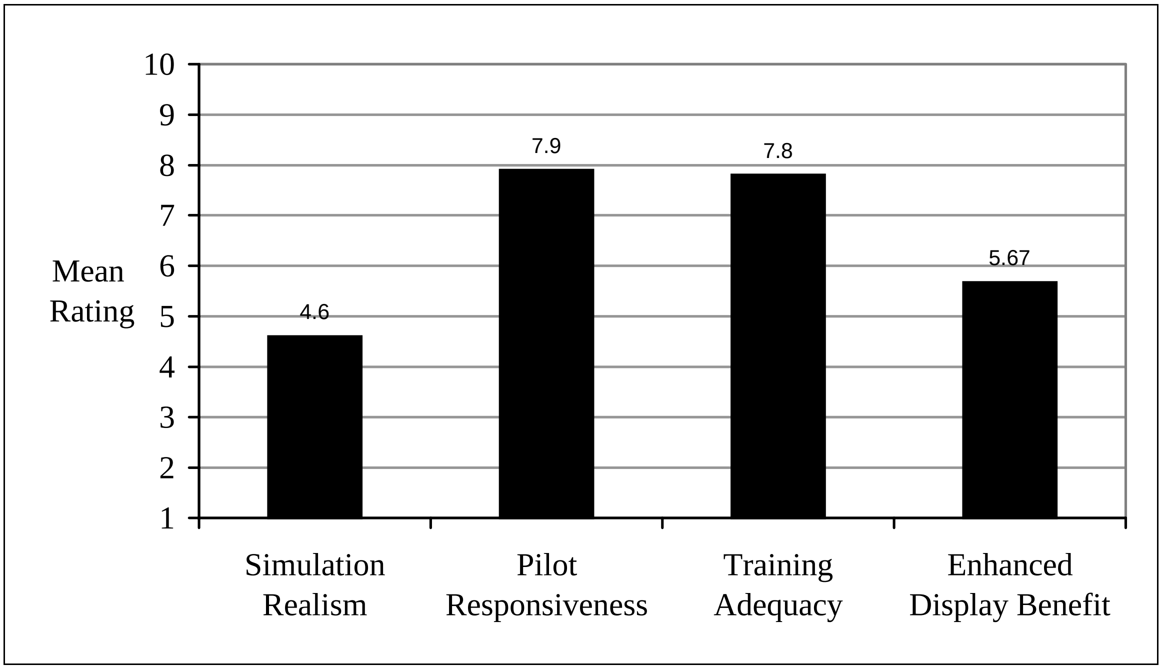

Figure 10. Mean rating of simulation characteristics.

3.5 Comparison of Situation Awareness Measures

A secondary objective of this research was to assess the utility of the different measures of SA that were included in the study. Both the SAGAT measure and the SART measure previously have received considerable work regarding validity. The researchers did not validate the on-line SA probes for SA measurement. The researchers carried out several analyses in order to examine the validity and comparability of these measures of SA.

3.5.1 Comparison of the On-line SA Probe with the SAGAT and Workload Measures

One question regarding the on-line SA probes concerns the degree to which they actually measure the participant SA, an internal representation of the ongoing situation. As an alternative hypothesis, they may actually measure workload, much as a secondary workload technique does. A secondary workload measure uses the time to respond to some stimulus as an indication of spare mental capacity. The on-line SA probe RT measure examined in this study may provide an indication of spare capacity (workload), or it may provide an indication of the degree to which needed information is actively present in working memory (an index of SA).

To examine these issues, the researchers performed a stepwise regression on SA probe RT. Independent variables were the SA Level of the probe, mean accuracy on the corresponding SAGAT query during the trial, the ATWIT rating taken within the same 1-minute interval, and the RT for that ATWIT rating. The regression model produced included SA Level and ATWIT RT and was statistically significant[$F(2, 477) = 3.48$, $p = .032$] although the multiple correlation was low ($R^2 = .014$). Neither mean accuracy on the corresponding SAGAT query nor the ATWIT rating were significantly related to SA probe RT. SA Probe RT was most related to the SA Level of the probe, $R^2 = .009$, followed by ATWIT RT, $R^2 = .006$, although these correlations are very low.

3.5.2 Predictive Validity of Situation Awareness and Workload Measures

In previous research, Durso et al. (1998) examined the use of an on-line SA probe in an ATC simulation. They separated RT to probes about the present situation (corresponding to our Level 1 SA probes) and RT to probes about the future situation (corresponding to our Level 2/3 SA probes) into two combined measures. They found that these measures were somewhat predictive of performance in their ATC simulation. To duplicate this approach, the researchers formed two combined scores for the SA probes. Similarly, the researchers developed combined SAGAT accuracy scores that separated queries about the present situation and the future situation. The SA measures included in the analysis were

a. SA Probe RT,
   1. SA Level 1 Probe RT, and
   2. SA Level 2/3 Probe RT.
b. SAGAT Queries,
   SAGAT Level 1-query accuracy (for aircraft presently under control),
   SAGAT Level 1-query accuracy (for aircraft to be under control in near future), and
   SAGAT Level 2 & 3 query accuracy.
c. SART, and
   supply,
   demand,
   understanding, and
   overall SART.
d. SME SA rating.

The workload measures included in the analysis were

a. Participant ratings, and
   1. NASA-TLX rating, and
   2. ATWIT.
      a) Mean ATWIT rating,
      b) Peak ATWIT rating, and

c) ATWIT RT.

b. SME ratings.

SME workload rating, and
SME NASA-TLX rating.

As indices of performance, the researchers selected the SMEs overall rating of performance and the mean of all their subjective ratings. Whereas the Durso et al. (1998) study examined the number of remaining actions at the end of the scenario (as an indication of how well the controller was keeping up with the job) as a measure of performance, no such measure was available in the present study. Instead, the researchers selected the number of flights completed (which was sensitive to the display manipulation), the number of conflicts in the trial, and duration of conflicts as general measures of performance.

The researchers submitted the SA and workload variables to stepwise regressions for each of these five dependant variables. For each resultant model, we submitted the variables included to independent regressions to verify their unique contribution to the model. The researchers will only discuss measures that showed a significant independent contribution to each model. A summary of these findings is in Table 6.

Table 6. Summary of Regression Models for Predictability of Performance from SA Measures

Model	***F*-Value**	**Significance**	**Adjusted R^2**
SME Overall Performance Rating			
SME SA Rating	$F(1,52) = 152.00$	$p <. 001$	$R^2 = .727$
SME Mean Performance Rating			
SME SA Rating	$F(1,58) = 287.45$	$p < .001$	$R^2 = .832$
SART- Understanding	$F(1,57) = 7.93$	$p =. 007$	$R^2 = .122$
Overall model	$F(2,56) = 147.26$	$p < .001$	$R^2 = .840$
Number of Flights Completed			
(none)			
Number of Conflicts			
SA Level 2/3 Probe RT	$F(1,57) = 6.14$	$p = .016$	$R^2 = .097$
SME SA Rating	$F(1,57) = 5.49$	$p = .023$	$R^2 = .088$
Overall model	$F(2,56) = 5.20$	$p = .008$	$R^2 = .157$
Duration of Conflicts			
SA Level 2/3 Probe RT	$F(1,57) = 5.23$	$p = .026$	$R^2 = .084$
SART – Demand	$F(1,56) = 4.26$	$p = .044$	$R^2 = .071$
SME SA Rating	$F(1,57) = 5.23$	$p = .023$	$R^2 = .084$
Overall model	$F(3,54) = 6.41$	$p = .001$	$R^2 = .263$

Only one variable, the SME rating of SA, significantly relates to the SME rating of overall performance, accounting for 72.7% of the variance. Similarly, SME ratings of SA highly correlate with the mean SME rating of performance, accounting for 83.2% of the variance in this measure. The understanding component of SART, another subjective measure, accounted for 12.2% of the variance in the mean SME performance ratings, although the overall model only accounted for 84% of the variance. This indicated some correlation between these two measures. The SME subjective ratings of performance highly correlate with their own subjective ratings of SA and, to a lesser degree, the participant subjective ratings of SA. All relationships are in the expected direction with the performance ratings positively correlated with the SA ratings.

The objective measures of performance show a considerably different picture. None of the SA or workload measures significantly correlate with the number of flights completed during the trial. The RT for the Level 2/3 SA probes accounts for 9.7% of the variance in the number of conflicts experienced during the trial. SME SA ratings account for 8.8% of the variance, with 15.7% of the variance accounted for overall. Whereas higher SME SA ratings were associated with fewer conflicts, as expected, slower RT to the Level 2/3 SA probes was associated with fewer conflicts. This was not in the expected direction. Controllers experiencing more conflicts answered the SA probes more quickly, and controllers experiencing fewer conflicts answered the probes more slowly.

The duration of conflicts similarly relates to Level 2/3 SA probe RT with 8.4% of the variance and SME SA ratings also with 8.4% of the variance. Similarly, higher SME SA ratings relate to shorter duration of conflicts. However, slower RT to Level 2 SA probes was associated with fewer conflicts. Additionally, the SART demand rating accounted for 7.1% of the variance, with higher demand ratings associated with longer duration conflicts. This led to a model that accounted for 26.3% of the overall variance in conflict duration.

This analysis did not show that the combined SAGAT measures significantly predict performance. Because the researchers found the scores provided by the individual SAGAT queries to be relatively independent, this apparent lack of sensitivity may be that the combined scores provide an inappropriate analysis of SAGAT data (Endsley, 1995b).[1]

We, therefore, conducted a second analysis of the SAGAT data. The researchers submitted the mean score for each trial for each SAGAT query to a stepwise regression procedure for each of the performance measures used in the study. Results show that SAGAT accounts for some of the variance in the objective performance variables (Table 7). The relationship to subjective performance ratings was very weak, however. In examining the objective measures of performance, the researchers predicted the number of conflicts by a combination of knowledge of vertical change (climbing, descending, or level), aircraft type, level of control (in sector control, handed-off, or coming into sector), and aircraft separation and advisory reception (28.7% of the variance). Similarly, we predicted the duration of conflicts by knowledge of vertical change,

[1] SAGAT queries are usually scored and analyzed individually. SA may vary considerably from item to item and may reflect changes in SA due to operator attention and display manipulations.

Table 7. Summary of Regression Models for Predictability of Performance from SAGAT

Model	***F*-Value**	**Significance**	**Adjusted R^2**
SME Overall Performance Rating	$F(1,57) = 4.79$	$p = .033$	$R^2 = .078$
Aircraft with Advisories			
SME Mean Performance Rating		not significant	
Number of Conflicts	$F(5,51) = 4.10$	$p = .003$	$R^2 = .287$
Vertical Change Type Level of Control Aircraft Separation Advisory Reception			
Duration of Conflicts	$F(4,55) = 3.03$	$p = .025$	$R^2 = .180$
Vertical Change Level of Control Aircraft Separation Aircraft with Advisories			
Number of Flights Handled		not significant	
Duration of Flights Handled	$F(3,52) = 5.29$	$p = .003$	$R^2 = .234$
Location of Sector Aircraft Speed Special Airspace Separation			
Distance Flown	$F(3,49) = 5.22$	$p = .003$	$R^2 = .242$
Location of Sector Aircraft Advisory Reception Special Airspace Separation			
Number of Completed Flight		not significant	
Cumulative Average Aircraft Density	$F(3,55) = 5.11$	$p = .003$	$R^2 = .218$
Altitude Aircraft Separation Aircraft in Communication			
Number of Heading Changes	$F(5,51) = 2.64$	$p = .035$	$R^2 =. 204$
Callsign Number Altitude Vertical Change			
Number of Altitude Changes	$F(3,53) = 4.12$	$p = .011$	$R^2 = .189$
Callsign Number Advisory Conformance Weather Impact			
Number of Transmissions	$F(3,49) = 4.025$	$p = .012$	$R^2 = .198$
Next Sector Advisory Reception Special Airspace Separation			
Duration of Transmissions	$F(4,48) = 6.187$	$p <.001$	$R^2 = .340$
Location of Sector Aircraft Location of All Aircraft Advisory Reception Special Airspace Separation			

level of control, aircraft separation, and aircraft with advisories outstanding (18% of the variance). For conflicts, keeping up with aircraft vertical maneuvers, separation with other aircraft, and status with regard to advisories appeared to be the most important.

SAGAT did not predict the number of completed flights and the number of flights handled. Knowledge of aircraft location, speed, and special airspace separation predicted the duration of the flights handled (23.4% of the variance). Knowledge of aircraft location, advisory reception, and special airspace separation predicted distance flown in the sector (24.2% of the variance). Knowledge of special airspace separation and aircraft location appears to be important for these two efficiency variables, whereas knowledge of speed was more closely related to how much time the aircraft spent in the sector.

The researchers predicted the number of heading changes by knowledge of callsign, altitude, and vertical change (20.9% of the variance). The researchers predicted the number of altitude changes by knowledge of callsign number, advisory conformance, and weather impact (18.9% of the variance). It would appear that decisions about whether to make a heading change may be closely related to knowledge regarding aircraft altitude and vertical maneuvering. Decisions to make altitude changes, however, more closely link to the need for changes due to inclement weather problems and whether aircraft are conforming to advisories. It is also possible that controllers pay more attention to advisory conformance when they give an altitude change. They generally attend to callsign in relation to issuing such changes.

They predicted the number of transmissions by knowledge of the aircraft next sector, advisory reception, and special airspace separation (19.8% of the variance). They predicted the duration of transmissions by location of sector aircraft, location of all aircraft (including those handed-off or not yet in sector), advisory reception, and special airspace separation (34% of the variance). Again, knowledge of special airspace separation significantly related to the efficiency measures. Attending to an aircraft present location and next sector also related to the need to communicate with the aircraft and other sectors.

3.5.3 Inter-Correlation of Situation Awareness Measures

Finally, the researchers made an analysis of the degree to which the SA measures examined in this study relate with each other. The researchers submitted the SA measures to a Pearson's correlation matrix. Tables 8 and 9 depict those correlations reaching a .05 level of significance.

Table 8. Inter-Correlation of SART Measures

	Supply	Demand	Understanding
SART	.539	-.523	.751
Supply		.310	.562
Demand			----

Table 9. Inter-Correlation of Other SA Measures

	SAGAT Level 1 (future)	SAGAT Level 2/3	SME SA rating
SART	.306	----	----
Understanding	----	-.326	.295
Supply	.258	-.292	----
SA Level 1 Probe RT	-.267	----	----

The SART overall score correlated highly with its three subcomponents, which also correlated with each other. Surprisingly, the researchers found a weak positive correlation between supply and demand, indicating that even when participants believed the demands were higher, they also believed their own supply of resources was higher. While understanding and supply correlated positively, the researchers found no significant correlation between understanding and demand.

The researchers found only moderate correlations between the other SA measures. The Level 1 SAGAT query score pertaining to future aircraft correlated moderately with the SART overall rating, the supply subcomponent, and the SA Level 1 probe RT. When participant knowledge of the status of incoming aircraft was higher, they tended to rate the supply of resources as higher, and the overall SART score was higher. They also responded faster to the SA Level 1 probes. These relationships are in the expected direction.

Scores on the Level 2/3 SAGAT queries pertaining to comprehension and projection of the aircraft correlate negatively with SART ratings of both understanding and supply. When participants show better knowledge in terms of comprehension and projection, they tend to subjectively rate understanding as lower and supply of resources as lower on the SART battery. The negative relationship between the SART understanding rating and the SAGAT comprehension/projection score was not expected. Finally, the SART understanding rating positively correlated with the SME SA rating.

3.5.4 Subjective Evaluation of the Measures

In the post-experiment questionnaire, the participants subjectively evaluated the three measures that were interjected during the trial, SAGAT, the on-line SA probe, and ATWIT for their subjective degree of interference. The on-line SA probe and ATWIT received low mean ratings (Figure 11). SAGAT, which required freezes in the simulation, received a higher overall rating of 5.2, showing a moderate level of subjective interference.

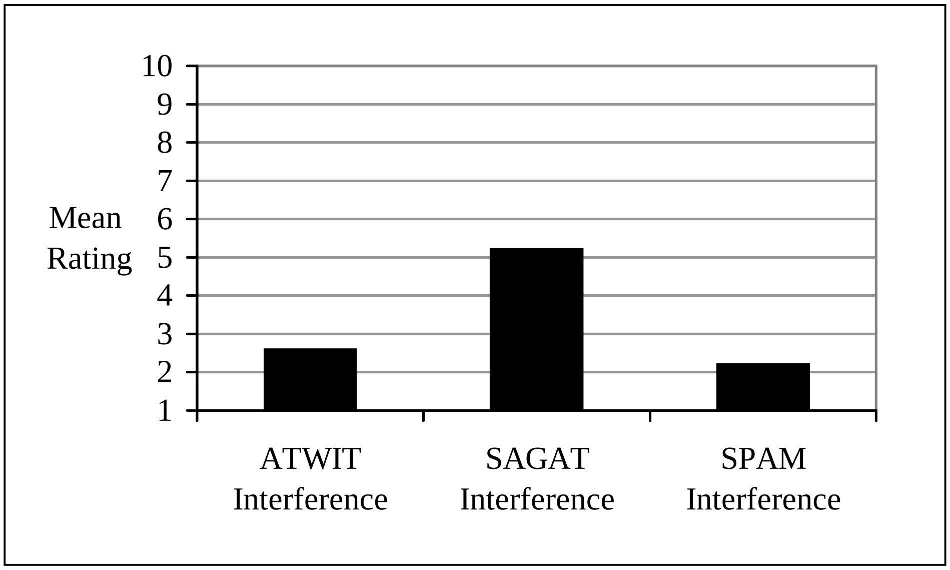

Figure 11. Mean rating of measure interference.

4. Discussion

The following subsections discuss the effectiveness of the enhanced display, experience with the self-separation concept employed in the study, and the utility and validity of the SA measures examined.

4.1 Enhanced Display

The introduction of the enhanced display provided some help for controllers in dealing with air traffic that are operating under self-separation. Controllers completed more flights with the enhanced display and showed a trend at being more able to correct their own errors.

The majority of the workload measures did not reflect these improvements , however. The fact that the mental workload component was slightly higher in the enhanced display condition was most likely reflective of the extra workload needed to monitor a second separate display. Controllers commented that placing the display window in a convenient spot without blocking radar data was a problem and having to locate the corresponding data block took extra effort. Many of the controllers commented that they would have found an integrated display to be more desirable.

Controller situation awareness also showed improvement. The extra information for transitionary aircraft showed a three-fold increase in knowledge of aircraft conformance to advisories. Endsley and Selcon (1997) previously found a reduction in keeping up with aircraft conformance under free flight conditions. The enhanced display appears to have been effective in combating this problem, most likely contributing to the trend towards savings in controller transmissions and reduction in heading changes.

Overall, the researchers recommend that the enhanced display receive further exploration. Efforts should be made to integrate the information with the controller radar picture, however, in order to reduce the workload associated with the display and potentially improve its utility even further.

4.2 Self-Separation Operational Concept

The self-separation operational concept employed in this study was very unique and a different experience for the controllers participating in the study. It represented a radical departure from the current operations. As such, the controllers viewed the simulations as unrealistic, although they rated training and pilot responsiveness as being reasonably good. In subjective comments, they found such a concept hard to imagine and were skeptical of it working in air traffic operations.

The study employed hypothetical on-board systems to simulate such an operational concept. No real assessment of either the desirability or feasibility of the self-separation concept was possible or intended with this study. Instead, it represented a worst case scenario in which controllers would assume a passive role in monitoring traffic. The pilots could choose whether to follow the controller recommendations on separation. The controllers found this very frustrating and contrary to how they normally controlled traffic. They normally would employ separation tactics far in advance, but, with this concept, they were forced to watch as aircraft came together and wonder if the aircraft would heed advisories to separate. They found this process involved more work than if they had been able to correct the situation and move on.

Whether or not ATC operations ever evolve towards such an operational concept is open to much debate. If it does, controllers and pilots will need additional aids such as the one employed in this study. It is recommended that the enhanced display, suitably modified, explored in this study also be examined within the context of current operational concept. The additional information it provides may be useful under these conditions as well.

4.3 Utility of Situation Awareness Measures

A secondary objective of this study was to examine the sensitivity, validity, and utility of the SA measures.

4.3.1 Sensitivity of the Measures

While SART and the on-line SA probe did not show a significant difference between conditions, it is difficult to say whether this reflected a problem with sensitivity or with the display manipulation in the study. A bigger question remains regarding the validity of the measures. Are they really measuring SA or are they reflecting something else? SAGAT and SART are the most widely used and validated measures of those included in the study (Garland & Endsley, 1995). The on-line probe was a relatively new measure of SA.

4.3.2 On-line Situation Awareness Probe Reaction Time

Endsley (1996) stated that interjecting questions during the performance of a task to measure SA was potentially intrusive. It added an additional task and potentially changed SA itself by redirecting attention to the information requested. We found no evidence to support the concern of intrusiveness in this study. The controllers subjectively rated its intrusiveness as low and reported that they often responded to such questions from other controllers. It is not possible to assess whether the measure actually led to changes in SA based on this study. To examine this possibility, the researchers recommend further studies to compare performance with such SA probes to performance in trials where the probes were not present.

Further examination of the on-line SA probe measure revealed concerns that are more fundamental. When the researchers compared the time to respond to each on-line SA probe to mean accuracy on the SAGAT query for the same information in that trial, the researchers found no relationship. A low but significant correlation exists between SA probe RT and RT to the ATWIT tone that occurred within 1 minute of the probe, indicating a weak relationship with workload. It may be that these findings reflected the temporal distance between when the SA probes occurred and when the SAGAT and ATWIT ratings occurred. That is, SA can be dynamic in nature, and a one-to-one comparison of these measures taken at different points in time may not provide a fair comparison. This possibility needs further exploration with more closely timed SA and workload measures.

The interpretation of the measure was also an issue. In examining the relation between the Level 2 & 3 SA probe RT measure and the number and duration of conflicts, the researchers observed a relationship in the opposite direction of that expected. The researchers cannot explain why slower responses to the comprehension and projection questions related to better performance. It is possible that controllers felt more rushed when they were having problems and, therefore, responded more quickly in these situations. If this is the case, however, it leaves interpretation of the on-line SA probe measure open to question.

A direct comparison was not possible. However, it should be noted that Durso et al. (1998) also found only the future-oriented SA probe reaction times to be related to the objective performance measure in their study. That relationship was opposite from the present study, however, and accounted for 13% of the variance. One possible difference is the self-separation operational concept that may have limited the degree to which SA could have an impact on performance. The researchers recommend that the on-line SA probe RT measure receive further testing within the context of the current operational concept.

4.3.3 Subjective Situation Awareness Measures

In conflict with the Durso et al. (1998) study, the researchers did not find a significant relationship between the Level 1 SA probe RT and the SME performance ratings. Instead, the SME SA rating was highly correlated with the SME performance ratings. Endsley (1996) criticized the lack of independence of subjective ratings of SA and subjective ratings of performance. Two of the performance measures examined also weakly correlated with SME SA ratings, however, indicating they probably considered such information in their ratings.

The SART measure performed reasonably well in the study, although the overall SART rating was not sensitive to the display manipulation. The measure possessed a fair degree of internal consistency, with the components inter-correlated with each other. The understanding component also correlated weakly with the SME performance rating, and the demand component was weakly correlated with the duration of conflicts.

The overall SART rating and the supply component correlated somewhat with the Level 1 SAGAT probes for future aircraft. This conflicted with a study by Endsley, Selcon, Hardiman, and Croft (1998) that found no relationship between SAGAT and SART in an aircraft task. The fact that the SART understanding and supply component were weakly correlated with the SAGAT comprehension and projection queries was somewhat perplexing. The researchers do not know exactly what aspects of SA people take into account when making subjective ratings, and this needs further exploration.

4.3.4 Situation Awareness Global Assessment Technique

The combined SAGAT measures did not relate to any of the performance measures examined here. This reflected the rather weak nature of the combined SAGAT scores. Little evidence exists supporting such a combination of queries. (The researchers made this calculation in order to provide direct comparison to the Durso et. al study.) When the individual SAGAT queries were used to predict performance, significant relationships resulted for most of the objective performance measures included in the test. These factors related in logical ways to each performance measure. This analysis would appear to support some predictive validity of the SAGAT technique.

SAGAT scores were sensitive to the display manipulation in this study and in previous studies (Endsley & Rodgers, 1996; Endsley & Selcon, 1997). In particular, they provided a clear indication of in what way the enhanced display was supportive of SA, providing good diagnosticity.

Controllers indicated that SAGAT was moderately intrusive. The on-line SA probes timing occurred so that the questions were not presented until after the controller had finished the present verbal communication. The SAGAT freezes, however, were implemented at the designated times, regardless of what was happening in the simulation. It may be desirable to deconflict more carefully the SAGAT freezes with controller actions so that they occur at times that are more desirable. Whereas previous studies have found no indication that the SAGAT freezes do in fact interfere with performance (Endsley, 1995a, b), these studies were done in aircraft simulations and probably should be repeated within the context of ATC simulations. The researchers recommend exploring the fact that the controllers expressed some concern about the artificiality of the freezes in order to weigh the costs of inserting such freezes against the benefit of the data provided by SAGAT.

4.3.5 Overall

This study represented the opportunity to examine the utility of various SA measures. The researchers did not specifically design it to validate any of these measures, however, and several factors limit these comparisons. First, the researchers do not know to what degree the self-separation concept may have interfered with the relationship between the SA measures and the performance measures examined. The researchers recommend examining the measures more carefully within the context of a normal operational concept.

Secondly, it is possible that the measures interfered with each other by virtue of occurring in the same trials. That is, it is possible that directing attention towards certain information with the on-line probes could alter SA, affecting the SAGAT measure. No evidence of this was present, however, as represented by the poor correlation between these measures. Secondly, it was possible for the ATWIT tones, by nature of their close temporal sequence to the on-line SA probes, to have affected performance on the probe. An analysis of whether the on-line SA probe occurred immediately before or after the ATWIT tone was not significant, however, allaying this concern. Of a more general nature, the participants completed many measures both during and following each trial. This may have been annoying. The researchers recommend keeping the number of measures to some minimum during testing to avoid fatiguing or demotivating the participants.

5. Conclusion

The enhanced display showed some promise for improving SA in advanced air traffic operations. Future research should modify the display to integrate the information with the existing radar displays and should test it under other operational conditions. This study expanded the knowledge base on methods for measuring SA by providing a direct comparison of several SA and workload measures. The researchers recommend further validation of the measures examined for use in ATC simulations.

Advanced air traffic operational concepts such as self-separation are interesting ideas with the potential of dramatically changing the NAS. The degree to which the system can accommodate such concepts without compromising aircraft safety, however, depends on establishing appropriate support within the ATC system. This study contributed to that effort by providing objective data on the effects of an enhanced display on controller ability to maintain an accurate and complete picture of the traffic situation. This is required in order to provide needed monitoring and separation support functions. Concepts such as self-separation may or may not happen, but the results of this study may also be generalizable to other areas. The enhanced display may also be helpful for promoting SA and facilitating performance under other operational rules.

References

ATCoach (Version 7.0) [Computer Software]. (1996). Lexington, MA: UFA, Inc.

Durso, F. T., Hackworth, C. A., Truitt, T. R., Crutchfield, J., Nikolic, D., & Manning, C. A. (1998). Situation awareness as a predictor of performance for en route air traffic controllers. *Air Traffic Control Quarterly, 6*(1), 1-20.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. *Proceedings of the Human Factors Society 32nd Annual Meeting* (pp. 97-101). Santa Monica, CA: Human Factors Society.

Endsley, M. R. (1995a). Direct measurement of situation awareness in simulations of dynamic systems: Validity and use of SAGAT. In D. J. Garland & M. R. Endsley (Eds.), *Experimental analysis and measurement of situation awareness* (pp. 107-113). Daytona Beach, FL: Embry-Riddle University.

Endsley, M. R. (1995b). Measurement of situation awareness in dynamic systems. *Human Factors, 37*(1), 65-84.

Endsley, M. R. (1996). Situation awareness measurement in test and evaluation. In T. G. O'Brien & S. G. Charlton (Eds.), *Handbook of human factors testing & evaluation* (pp. 159-180). Mahwah, NJ: Lawrence Erlbaum.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors, 37*(2), 381-394.

Endsley, M. R., Mogford, R., Allendoerfer, K., Snyder, M. D., & Stein, E. S. (1997). *Effect of free flight conditions on controller performance, workload and situation awareness: A preliminary investigation of changes in locus of control using existing technology* (DOT/FAA/CT-TN 97/12). Atlantic City International Airport, NJ: Federal Aviation Administration William J. Hughes Technical Center.

Endsley, M. R., & Rodgers, M. D. (1994). *Situation Awareness Global Assessment Technique (SAGAT): En route air traffic control version user's guide* (Draft). Lubbock: Texas Tech University.

Endsley, M. R., & Rodgers, M. D. (1996). Attention distribution and situation awareness in air traffic control. In *Proceedings of the 40th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 82-85). Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M. R., & Selcon, S. J. (1997). Designing to aid decisions through situation awareness enhancement. In *Proceedings of the 2nd Symposium on Situation Awareness in Tactical Aircraft* (pp. 107-112). Patuxent River, MD: Naval Air Warfare Center.

Endsley, M. R., Selcon, S. J., Hardiman, T. D., & Croft, D. G. (1998). A comparative evaluation of SAGAT and SART for evaluations of situation awareness. In *Proceedings of the Human*

Factors and Ergonomics Society Annual Meeting (pp. 82-86). Santa Monica, CA: Human Factors and Ergonomics Society.

Garland, D. J., & Endsley, M. R. (Ed.). (1995). *Experimental analysis and measurement of situation awareness*. Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

Gutman, J. A., Stein, E., & Gromelski, S. (1995). *The influence of generic airspace on air traffic controller performance* (DOT/FAA/CT-TN95/38). Atlantic City International Airport: FAA William J. Hughes Technical Center.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North-Holland.

Sollenberger, R. L., Stein, E. S., & Gromelski, S. (1997). *The development and evaluation of a behaviorally based rating form for assessing air traffic controller performance* (DOT/FAA/CT-TN96/16). Atlantic City International Airport, NJ: Federal Aviation Administration 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.

Stein, E. S. (1991). Evaluating air traffic controller workload using real time person in the loop simulation. *Journal of Air Traffic Control*, *33*(4), 55-58.

Taylor, R. M. (1990). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In *Situational awareness in aerospace operations* (AGARD-CP-478) (pp. 3/1 - 3/17). Neuilly Sur Seine, France: NATO - AGARD.

Appendix A
Subject Matter Expert Subjective Rating Form
Air Traffic Control Evaluation for Self-Separation Operational Concepts

Instructions for questions 1-21

This form was designed to be used by instructor certified air traffic control specialist to evaluate the effectiveness of controllers working in simulation environments. Observers will rate the effectiveness of controllers in several different performance areas using the scale show below. When making your ratings, please try to use the entire scale range as much as possible. You are encouraged to write down observations and you may make preliminary ratings during the course of the scenario. However, the researchers recommend that you wait until the scenario is finished before making your final ratings. The observations you make do not need to be restricted to the performance areas covered in this form ands may include other areas that you think are important. Also, please write down any comments that may improve this evaluation form. Your identity will remain anonymous, so do not write your name on the form.

Rating	Label Description
1	Controller demonstrated extremely poor judgment in making intervention decisions and very frequently made errors
2	Controller demonstrated poor judgment in making some intervention decisions and occasionally made errors
3	Controller make questionable decisions using poor intervention techniques which led to restricting the normal traffic flow
4	Controller demonstrated the ability to keep aircraft separated but used spacing and separation criteria which was excessive
5	Controller demonstrated adequate judgment in making intervention decisions
6	Controller demonstrated good judgment in making intervention decisions using efficient control techniques
7	Controller frequently demonstrated excellent judgment in making intervention decisions using extremely good control techniques
8	Controller always demonstrated excellent judgment in making even the most difficult intervention decisions while using outstanding control techniques

Maintaining Safe and Efficient Traffic Flow

1. Maintaining Separation and Resolving Potential Conflicts — 1 2 3 4 5 6 7 8
   - using interventions that maintain safe aircraft separation
   - detecting and resolving impending conflicts early

   Comments: ______________________________

2. Using Separation Interventions Effectively — 1 2 3 4 5 6 7 8
   - providing accurate navigational assistance to pilots
   - avoiding interventions that result in the need for additional instructions to handle aircraft completely
   - avoiding excessive interventions or over-controlling

   Comments: ______________________________

3. Overall Safe and Efficient Traffic Flow Scale Rating — 1 2 3 4 5 6 7 8

   Comments: ______________________________

Maintaining Attention and Situation Awareness

4. Maintaining Awareness of Aircraft Positions — 1 2 3 4 5 6 7 8
   - avoiding fixation on one area of the radar scope when other areas need attention
   - using scanning patterns that monitor all aircraft on the radar scope

   Comments: ______________________________

5. Identifying Traffic Conflict Problems in a Timely Manner — 1 2 3 4 5 6 7 8
   - keeping up with traffic trajectories
   - projecting separation problems in a timely manner
   - avoiding excessive interventions or over-controlling

   Comments: ______________________________

6. Correcting Own Errors in a Timely Manner — 1 2 3 4 5 6 7 8

   Comments: ______________________________

7. Overall Attention and Situation Awareness Scale Rating — 1 2 3 4 5 6 7 8

   Comments: ______________________________

Prioritizing

8. Taking Actions in an Appropriate Order of Importance - resolving situations that need immediate attention before handling low priority tasks - issuing interventions in a prioritized, structured, and timely manner **Comments:** ______________________	1	2	3	4	5	6	7	8
9. Handling Tasks for Several Aircraft - shifting tasks between aircraft - avoiding delays in communications while thinking or planning actions **Comments:** ______________________	1	2	3	4	5	6	7	8
10. Keeping Datablocks Up-to-date - updating datablocks accurately while talking or performing other tasks - keeping datablocks updated in a timely manner **Comments:** ______________________	1	2	3	4	5	6	7	8
11. Overall Prioritizing Scale Rating **Comments:** ______________________	1	2	3	4	5	6	7	8

Providing Control Information

12. Providing Essential Air Traffic Control Information - providing mandatory services and advisories to pilots in a timely manner - exchanging essential information **Comments:** ______________________	1	2	3	4	5	6	7	8
13. Providing Additional Air Traffic Control Information - providing additional services when workload is not a factor - exchanging additional information **Comments:** ______________________	1	2	3	4	5	6	7	8
14. Overall Providing Control Information Scale Rating **Comments:** ______________________	1	2	3	4	5	6	7	8

Technical Knowledge

15. Showing Knowledge of LOAs and SOPs — 1 2 3 4 5 6 7 8
    - controlling traffic as depicted in current LOAs and SOPs
    - performing hand-off procedures correctly

    Comments: ______________________________

16. Showing Knowledge of Aircraft Capabilities and Limitations — 1 2 3 4 5 6 7 8
    - avoiding advisories that are beyond aircraft performance parameters
    - recognizing the need for speed restrictions and wake turbulence separation

    Comments: ______________________________

17. Overall Technical Knowledge Scale Rating — 1 2 3 4 5 6 7 8

    Comments: ______________________________

Communicating

18. Using Proper Phraseology — 1 2 3 4 5 6 7 8
    - using words and phrases specified in ATP 7110.65
    - using ATP phraseology that is appropriate for the situation
    - avoiding the use of excessive verbiage

    Comments: ______________________________

19. Communicating Clearly and Efficiently — 1 2 3 4 5 6 7 8
    - speaking at the proper volume and rate for pilots to understand
    - speaking fluently while scanning or performing other tasks
    - communication delivery is complete, correct and timely
    - providing complete information in each communication

    Comments: ______________________________

20. Listening for Pilot Readbacks and Requests — 1 2 3 4 5 6 7 8
    - correcting pilot readback errors
    - processing pilot requests correctly in a timely manner

    Comments: ______________________________

21. Overall Communicating Scale Rating — 1 2 3 4 5 6 7 8

    Comments: ______________________________

Instructions for questions 22-29

The following questions have as scale ranging from 1 to 10. Where 1 represents “extremely low”, “extremely infrequent”, “strongly disagree”, etc. and 10 represents the other extreme of the spectrum.

These questions are the same as the researchers have asked the controller after the scenario. The researchers would like you to give us your impression of how these questions will be rated by the controller.

22.	Please circle the number below that best describes **how hard the controller was working** during this scenario. **Comments:** ______	not hard	1	2	3	4	5	6	7	8	9	10	extremely hard

23.	Please circle the number that best describes **how well the controller managed traffic** during this scenario **Comments:** ______	extremely poor	1	2	3	4	5	6	7	8	9	10	extremely well

NASA TLX

24.	Please circle the number that best describes the **mental demand** during this scenario.	extremely low	1	2	3	4	5	6	7	8	9	10	extremely high
25.	Please circle the number that best describes the **physical demand** during this scenario.	extremely low	1	2	3	4	5	6	7	8	9	10	extremely high
26.	Please circle the number that best describes the **temporal demand** during this scenario.	extremely low	1	2	3	4	5	6	7	8	9	10	extremely high
27.	Please circle the number that best describes the overall **performance** during this scenario.	extremely low	1	2	3	4	5	6	7	8	9	10	extremely high
28.	Please circle the number that best describes the **effort** during this scenario.	extremely low	1	2	3	4	5	6	7	8	9	10	extremely high
29.	Please circle the number that best describes the level of **frustration** during this scenario.	extremely low	1	2	3	4	5	6	7	8	9	10	extremely high

Appendix B
Situation Awareness Global Assessment Technique Queries

1. Enter the location of all aircraft (on the provided sector map)
    aircraft in track control
    other aircraft in sector
    aircraft will be in track control in next 2 minutes
2. Enter aircraft callsign (for aircraft highlighted of those entered in query 1)
3. Enter aircraft altitude (for aircraft highlighted of those entered in query 1)
4. Enter aircraft groundspeed (for aircraft highlighted of those entered in query 1)
5. Enter aircraft heading (for aircraft highlighted of those entered in query 1)
6. Enter aircraft's next sector (for aircraft highlighted of those entered in query 1)
    A B C D E
7. Enter aircraft's current direction of change in each column (for aircraft highlighted of those entered in query 1)

<u>Altitude change</u>	<u>Turn</u>
climbing	right turn
descending	left turn
level	straight

8. Enter the aircraft type (for aircraft highlighted of those entered in query 1)

9. Which pairs of aircraft have lost or will lose separation if they stay on their current (intended) courses?

10. Which aircraft have been issued advisories for situations which have not been resolved?

11. Did the aircraft receive its advisory correctly? (for each of those entered in query 11)

12. Which aircraft are currently conforming to their advisories? (for each of those entered in query 11)

13. Which aircraft must be handed off to another sector/facility within the next 2 minutes?

14. Enter the aircraft which are not in communication with you.

15. Enter the aircraft that will violate special airspace separation standards if they stay on their current (intended) paths.

16. Which aircraft are weather currently an impact on or will be an impact on in the next 5 minutes along their current course?

Appendix C

10-Dimensional Situational Awareness Rating Technique (SART)

Please answer these questions with regard to the traffic situations presented in the scenario.

Instability of Situation

How changeable is the situation? Is the situation highly unstable and likely to change suddenly (high), or is it very stable and straight forward (low)?

Low ├──────────────────────────┤ High

Complexity of Situation

How complicated is the situation? Is it complex with many interrelated components (high) or is it simple and straightforward (low)?

Low ├──────────────────────────┤ High

Variability of Situation

How many variables are changing in the situation? Are there are large number of factors varying (high) or are there very few variables changing (low)?

Low ├──────────────────────────┤ High

Arousal

How aroused are you in the situation? Are you alert and ready for activity (high) or do you have a low degree of alertness (low)?

Low ├──────────────────────────┤ High

Concentration of Attention

How much are you concentrating on the situation? Are you bringing all your thoughts to bear (high) or is your attention elsewhere (low)?

Low ├──────────────────────────┤ High

Division of Attention

How much is your attention divided in the situation? Are you concentrating on many aspects of the situation (high) or focussed on only one (low)?

Low |————————————————————| High

Spare Mental Capacity

How much mental capacity do you have to spare in the situation? Do you have sufficient to attend to many variables (high) or nothing to spare at all (low)?

Low |————————————————————| High

Information Quantity

How much information have you gained about the situation? Have you received and understood a great deal of knowledge (high) or very little (low)?

Low |————————————————————| High

Information Quality

How good is the information you have gained about the situation? Is the knowledge communicated very useful (high) or is it a new situation (low)?

Low |————————————————————| High

Familiarity with Situation

How familiar are you with the situation? Do you have a great deal of relevant experience (high) or is it a new situation (low)?

Low |————————————————————| High

Appendix D
Air Traffic Workload Input Technique (ATWIT)

Rate your Workload

1	2	3	4	5	6	7	8	9	10

Appendix E

NASA-Task Load Index (TLX)

Please answer these questions with regard to the preceding scenario.

Mental Demand
How much mental and perceptual activity is required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Is the task easy or demanding, simple or complex, exacting or forgiving?

Low High

Physical Demand
How much physical activity is required (e.g., pushing, turning, controlling, activating. etc.)? Is the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Low High

Temporal Demand
How much time pressure do you feel due to the rate or pace at which the tasks or task elements occurred? Is the pace slow and leisurely or rapid and frantic?

Low High

Performance
How successful do you think you are in accomplishing the goals of the task? How satisfied are you with your performance in accomplishing these goals?

Good Poor

Effort
How hard did you have to work (mentally and physically) to accomplish this level of performance?

Low High

Frustration
How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent do you feel in performing the task?

Low High

Appendix F

Post-Trial Participant Subjective Questionnaire

1. Please circle the number below that best describes **how hard you were working** during this scenario. **Comments:**	not hard	1	2	3	4	5	6	7	8	9	10	extremely hard
2. Please circle the number that best describes **how well you managed traffic** during this scenario **Comments:**	extremely poor	1	2	3	4	5	6	7	8	9	10	extremely well
3. Please circle the number that best describes **how difficult** this scenario was. **Comments:**	extremely difficult	1	2	3	4	5	6	7	8	9	10	extremely difficult
4. Do you have any other **comments about your experiences during the simulation**? **Comments:**												

Appendix G
Post-Experiment Participant Subjective Questionnaire

1. Please circle the number that best describes **how realistic the simulations** were **Comments:** ______	extremely unrealistic	1 2 3 4 5 6 7 8 9 10	extremely realistic
2. Please circle the number that best describes **how well the simulation-pilots responded** to your advisories in terms of traffic movement or radio communication. **Comments:** ______	extremely poor	1 2 3 4 5 6 7 8 9 10	extremely well
3. Please circle the number that best describes if the **hands-on training for each scenario was adequate**. **Comments:** ______	not adequate	1 2 3 4 5 6 7 8 9 10	adequate
4. Please circle the number that best describes how much the **ATWIT device interfered** with your performance. **Comments:** ______	No affect on Performance	1 2 3 4 5 6 7 8 9 10	Extreme affect on performance
5. Please circle the number that best describes how much the **SAGAT freezes** interfered with your performance. **Comments:** ______	No affect on Performance	1 2 3 4 5 6 7 8 9 10	Extreme affect on performance
6. Please circle the number that best describes how much the **Queries during the simulation** interfered with your performance. **Comments:** ______	No affect on Performance	1 2 3 4 5 6 7 8 9 10	Extreme affect on performance
7. Please circle the number that best describes how much the **Enhanced Display assisted** your performance. **Comments:** ______	No affect on Performance	1 2 3 4 5 6 7 8 9 10	Extreme affect on performance

8. Is there anything about the study that we should have asked or that you would like to comment about?

Comments:

Appendix H
SME SAGAT Data Evaluation Form

Subject _________ **Condition __________** **Scenario _________Trial ___________** **Stop number __________**

* if aircraft stays on current (intended) path

	Aircraft	Track Control	Vertical velocity	Turning	Next Sector	Sector airspace violation in next 2 min*	Not in comm with sector	Will violate SUA next 2 min*	Weather will impact in next 5 min*
1		my control in next 2 min other in sector	level climbing descending	straight left right					
2		my control in next 2 min other in sector	level climbing descending	straight left right					
3		my control in next 2 min other in sector	level climbing descending	straight left right					
4		my control in next 2 min other in sector	level climbing descending	straight left right					
5		my control in next 2 min other in sector	level climbing descending	straight left right					
6		my control in next 2 min other in sector	level climbing descending	straight left right					
7		my control in next 2 min other in sector	level climbing descending	straight left right					
8		my control in next 2 min other in sector	level climbing descending	straight left right					

Which pairs of aircraft have lost or will lose separation if they stay on their current

(intended) courses?

Which aircraft have been issued advisories for situations that are not yet resolved?

	Received correctly?	Conforming to advisory?
	Y / N	Y / N
	Y / N	Y / N
	Y / N	Y / N
	Y / N	Y / N
	Y / N	Y / N