ote technical note techi

Effects of Collocation and Reduced Lateral Separation Standards in the New York Integrated Control Complex

Todd R. Truitt, Ph.D.
D. Michael McAnulty, Ph.D.
Ben Willems, M.A.

March 2004
DOT/FAA/CT-TN04/08

U.S. Department of Transportation
Federal Aviation Administration

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

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/CT-TN04/08		

4. Title and Subtitle Effects of Collocation and Reduced Lateral Separation Standards in the New York Integrated Control Complex	**5. Report Date** March 2004 **6. Performing Organization Code**
7. Author(s) Todd R. Truitt, Ph.D., D. Michael McAnulty, Ph.D., and Ben Willems (ACB-220)	**8. Performing Organization Report No..** DOT/FAA/CT-TN04/08
9. Performing Organization Name and Address Federal Aviation Administration NAS Human Factors Group, ACB-220 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 Eastern Region Air Traffic Division Airspace Branch One Aviation Plaza Jamaica, New York 11434-4809	**13. Type of Report and Period Covered** Technical Note **14. Sponsoring Agency Code** AEA-520

15. Supplementary Notes

16. Abstract

The authors used a high fidelity, human-in-the-loop simulation in two experiments examining the concept of the New York Integrated Control Complex (NYICC). The first experiment examined the potential effects of collocation alone and collocation with expanded terminal separation standards on the arrival traffic flow into Newark. The second experiment examined the same effects on the departure traffic flows primarily from Newark International (EWR) and LaGuardia (LGA) airports. Each experiment compared three different conditions. In the Normal condition, the participants controlled traffic as they normally would, and a removable wall physically separated the terminal and en route sectors. During the Collocated condition, experimenters removed the wall, and the participants could engage in face-to-face communication and look at each other's radar displays. During the Terminalized condition, we collocated the sectors and reduced the separation standard for one or both en route sectors from 5 nm to 3 nm. We collected measures throughout the experiments including system performance, subjective ratings of workload, subject matter expert ratings of performance, communication behaviors, and participant opinion. Overall, both experiments provided support for the NYICC concept of operations.

17. Key Words	18. Distribution Statement
New York Integrated Control Complex Communication and Coordination Human-in-the-loop simulation Reduced separation Air Traffic Control Collocation	This report is approved for public release and is on file at the William J. Hughes Technical Center, Aviation Security Research and Development Library, Atlantic City International Airport, New Jersey, 08405. This document is also available to the U.S. public through the National Technical Information Service (NTIS), Springfield, Virginia, 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	192	

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

Acknowledgements

These experiments were made possible by funding from the Federal Aviation Administration's Terminal Business Unit (ATB-1) and the Eastern Region Airspace Branch (AEA-520). In this regard, we would like to thank Bill Voss and Mike Sammartino. We would also like to thank the members of the New York Airspace Redesign Team including Steve Kelley, Timon Kalpaxis, Robert Clark, Tom O'Neill, and Jim Shelton for their help in preparing the experiment, shakedown of equipment and procedures, and the scheduling of participants. Thanks to Loretta Martin, Air Traffic Manager, New York Terminal Radar Approach Control (TRACON), and Michael McCormick, Air Traffic Manager, New York Air Route Traffic Control Center (ARTCC), for providing staff and participants for the experiment. From the New York TRACON, we would like to especially acknowledge John Cirincione, William D'Alo, Phil Rodgers, William Russell, and Tom Zaccheo who participated in the shakedown runs and coordinated participants during the simulations. Also, William Deacy and Dave Wynkoop ran the shakedown exercises for departures. From the New York ARTCC, special thanks to Dan Frasier who participated in the shakedown runs and coordinated participants during the departure simulation. Also, George Acampora, John Higgins, Phil Sudano, and Greg Gilman ran the shakedown exercises for departures and Sal Messineo coordinated participants.

We give special recognition to John Goldman and Joe Toland of Northrop Grumman. John was our lead Air Traffic Control Subject Matter Expert (SME). He worked with the Airspace Redesign team to obtain information about the New York traffic flows and then built multiple 1 hr traffic scenarios that were extremely busy yet highly realistic. Prior to each simulation, he played a key role in exercising the scenarios to identify and remove any problems. He also coordinated with the simulation pilots to make sure their performance was realistic. He worked with the Target Generation Facility (TGF) personnel to make sure the radar targets were accurate and that aircraft performance was consistent with actual capabilities. He also worked with them to create a new simulation feature that would hold departures at specified airports. John also served as an over-the-shoulder evaluator of terminal controller performance during the simulations. After the simulations, he helped us interpret the findings and brief the sponsors. Finally, John provided great guidance and insight to our human factors personnel and the technical staff to help us do our jobs more effectively. Joe Toland, although new to the Research, Development, and Human Factors Laboratory (RDHFL) and the project, quickly became a valued contributor by working with the airspace group and our other staff members to develop and debug the multiple scenarios used in these projects. He also participated in the shakedown exercises to identify and remove any problems. He provided information to populate and prepare the flight progress strips for each scenario. During the simulation runs, Joe served as an over-the-shoulder evaluator of terminal controller performance. He also provided valuable information and insight about air traffic operations to the other members of our staff.

Special recognition also goes to Dr. Randy Sollenberger of ACB-220. Based on his extensive experience conducting high fidelity, human-in-the-loop simulations, Randy provided valuable advice to us on how to set up and run the simulations and on how to analyze the resulting data. He also developed condition labels for the departure simulations that were very beneficial in keeping the 10 participants working effectively. When we unexpectedly lost one of our staff members during the arrival simulation, he stepped in and collected the face-to-face communications data. Perhaps most important, Randy wrote a program to transfer the aircraft flight plans so they would print out in flight strip format, which enabled us to prepare the hundreds of flight strips needed for each scenario run.

Thanks to Robert Oliver, ACB-840, who led the computer programming team in developing the simulations, conducting shakedown runs, and running the test scenarios. Dan Johnson, Dario Rosilli, Joanne Flood, Boskey Patel, A.J. Mieskolainen, Joe Keen, Yev Tabekman, Jim Merel, and Ralph "Ed" Little of Titan Systems were also crucial to the software engineering effort. Albert Macias, Mary Delemarre, Lou Delemare, and Joe Diluzio of ACB-840, and Otto Smith, George Rowand, Tom D'Ottavi, Jr., and Steve Senesi of Titan Systems provided engineering support to set up the simulation hardware and audio-video recording and to monitor the networks and data collection process. Jim Valleley, ACB-840, provided assistance in locating and ordering equipment needed to set up and run the simulations. Thanks also to Russell Spadea, ATB-230, Robert Angelini, Raytheon, and Emmanuel Papadopoulos, DMS Tech, for loaning us two Standard Terminal Automation Replacement System (STARS) desktop keyboards and two desktop trackballs that we needed to run the simulations.

We also want to recognize the efforts of Al Rehmann and ACB-860 for coordinating the TGF and simulation pilot resources. Dan Warburton led the TGF team that provided development, runtime, aircraft performance modeling, and data collection and analysis support to the effort. Scott Doucett set up new simulation pilot workstations required to support the five radar sectors in the departure simulation. Samantha Fullerton, Bethany Leffler, Dana Whicker, Lonnie Souder, Stan Rimdzius, Mike Konyak, and Mike Capito from the TGF provided development, runtime, aircraft performance modeling, and data collection and analysis support to the effort.

Also from ACB-860, Mary Rozier-Wilkes, Lelia Dixon, Barbara Para, Leonora Richardson, Audrey Wilmer, Robert Engiles, Tru Hall, Bill Hickman, Dale Laudenslager, Bruce Slack, and John Szuba did an excellent job as simulation pilots during shakedown and the test runs. From Titan Systems, Cindy Hogan, Mary Ann Smith, Mary Lou Hagan, Joe Guida, Skip Bellino, Frank Johnson, and Shawn Sivak also did an excellent job as simulation pilots during shakedown and the test runs.

Thanks to Dr. Michele Heiney of Northrop Grumman who developed a data collection form and procedures to capture face-to-face and nonverbal communications during the Collocated and Terminalized conditions. During the simulations, she gave the participating controllers instructions on workload rating procedures prior to each scenario, collected face-to-face and nonverbal communications data, and monitored the total operation (e.g., the simulator system and TGF) to make sure it was performing properly. After each scenario, she administered the post-scenario questionnaires to the controllers and collected the SME ratings data. Finally, Michele participated in analyzing the communications data.

Thanks to Kevin Hallman of Northrop Grumman who set up and entered flight information into a data base, printed flight progress strips for all positions and aircraft, stuffed them into holders, and had them available in proper order at each position when needed by the controllers. After the simulations, he set up multiple databases in the proper format to conduct repeated measures analyses, entered all the subjective data collected, and participated in computing statistics. Thanks also to Mark Hale of Titan Systems who provided assistance in preparing the flight progress strips. Fa Jackson, Titan Systems, also assisted us in analyzing the data after the experiment.

Thanks to Phil Basset and David Conley of the Jacksonville ARTCC for serving as over-the-shoulder SME raters. Phil assisted during the arrival simulation and Dave assisted during the departure simulation. Richard Ridgway, Titan Systems, also worked with us during shakedown and served as an over-the-shoulder SME rater during the departure simulation.

Table of Contents

List of Illustrations

Tables Page

Executive Summary

The National Airspace System (NAS) is under increasing pressure due to the congested airspace in the Northeast United States. The Federal Aviation Administration (FAA) does not expect this congestion to resolve itself due to the prediction of increasing air traffic and the lack of plans to construct new airways. Furthermore, the FAA has not updated the airspace design and procedures that Certified Professional Controllers use quickly enough to keep pace with the increasing levels of air traffic. The New York Airspace Redesign Team has developed a concept of operations called the New York Integrated Control Complex (NYICC) to address the congested NAS. In addition to redesigning the structure of the airspace, there are two basic elements of the NYICC. First, it would collocate the New York terminal and en route facilities to improve communication and coordination between them. Second, it would expand the terminal airspace to reduce the number of transfer of control points and to move these points further from the airports to improve overall traffic flow. The expanded terminal airspace also would allow traffic sequencing, spacing, and holding to occur closer to the arrival airports.

The purpose of the present experiments was to test two components of the NYICC concept using a high fidelity, human-in-the-loop simulation. Engineering Research Psychologists conducted two experiments at the FAA William J. Hughes Technical Center Research, Development, and Human Factors Laboratory. The first experiment examined the potential effects of collocation alone and collocation along with expanded terminal separation standards on the arrival traffic flow into Newark International Airport (EWR). The second experiment examined the same effects on the departure traffic flows primarily from EWR and LaGuardia Airport (LGA), but included other airports within the same airspace. Each experiment examined and compared three different conditions. The Normal condition served as a baseline. In this condition, the participants controlled traffic as they normally would. A removable wall physically separated the terminal and en route sectors during the Normal condition. During the Collocated condition, we removed the wall separating the terminal and en route sectors, and the participants were allowed to engage in face-to-face (FTF) communication. The participants could also look at each others' radar displays during this condition. During the Terminalized condition, we collocated the terminal and en route sectors, and we reduced the separation standard for one or both en route sectors from 5 nautical miles (nm) to 3 nm. Throughout the experiments, we collected numerous measures including system performance, subjective ratings of workload, subject matter expert ratings of performance, communication behaviors, and participant opinion.

Overall, both experiments provided support for the NYICC concept of operations. In particular, the Terminalized condition, (i.e., collocation of the terminal and en route facilities, along with expanded terminal separation standards) provided the greatest benefits compared to the Normal condition. In the Terminalized condition, the participants were able to increase the number of arrivals they could provide and the number of departures they could accept. They reduced the number and duration of holds, the number and duration of departure stops, and the time between departures. The participants also were able to reduce the number of landline communications while compensating with the use of FTF communication. The participants capitalized further on their collocated situation by looking at each other's radar displays and, thereby, improving their situation awareness of the air traffic situation. The Terminalized condition caused some slight increases in the participants' taskload and workload; however, these increases did not exceed moderate levels. There were no negative effects on safety. The participants reported overall improvements in the ease of communication and coordination and in the use of airspace and traffic flow.

The experiment was limited in scope to the sectors, conditions, and air traffic scenarios that we used, but the overall result was positive for the NYICC concepts. We recommend further studies to examine the impact of the NYICC on other adjacent facilities such as air traffic control towers, ground control, adjacent air route traffic control centers, and traffic management units. Furthermore, researchers should further examine the concept of collocation to determine how best to arrange the terminal and en route sector positions to maximize the effectiveness of their communication and coordination. Finally, we anticipate that further benefits can be achieved through improved procedures and training to exploit the advantages of collocation and reduced separation standards.

FEDERAL AVIATION ADMINISTRATION
Sponsor Summary

From: Eastern Terminal Area
Date: Monday, March 01, 2004
Re: Effects of Collocation and Reduced Lateral Separation Standards in the New York Integrated Control Complex (NYICC)

Overview: The Eastern Region Air Traffic Division was congressionally mandated to undertake a program which relieves the congested airspace in the Northeast United States. During the redesign phase, the team noted improvements could be realized by collocating the New York TRACON (N90) and the New York ARTCC (ZNY), and applying terminal separation standards to the maximum extent possible.

Background: In an effort to determine whether the integration of N90 and ZNY would produce realized benefits in system improvements, the Human-in-the-Loop study was undertaken.

The study criteria examined the effects of collocation and the resultant terminalization of current enroute airspace. The experiment simulated both arrival and departure flows. The first element examined the potential effects of collocation alone and collocation along with expanded terminal separation standards on the arrival traffic flow into Newark. The second element examined the same effects on the departure traffic flows primarily from Newark and LaGuardia, but it also included other airports within the airspace.

Purpose/Goal: The goal was to validate the theory that collocation of the facilities and further terminalization of existing airspace (routes and flows) would produce tangible benefits related to the operation of the NAS, and subsequently pass these benefits on to the customer. The basis of this theory is devoid of any major airspace redesign initiatives and is based on existing routes and flows.

Outcome/Final Position: When accuracy of the HITL Draft Report was questioned, NASA was tasked to provide an independent analysis of the experiment.

NASA's initial concerns and comments were resolved by the study's authors. NASA's final recommendation strongly suggests that the study be taken in context as it only addresses some aspects of the NYICC. NASA believes that other studies, such as facility layout, airspace and procedural changes, safety and cost/benefit should be incorporated into the evaluation of the concept.

After a final review of the product, it is the sponsor's opinion that the study has shown scientific benefit to the collocation of the facilities and expanded terminalization of the airspace. However, further validation requires a scenario presenting the expanded terminalization of current enroute airspace without collocation of the facilities. It also requires the incorporation of those items identified by NASA.

1. INTRODUCTION

As the number of air carrier flights continues to climb, there is an ever-increasing effect on the efficiency of the National Airspace System (NAS). The northeast portion of the United States, in particular, is becoming congested by air traffic. According to the Federal Aviation Administration (FAA) Research and Development Strategy (2002a), the FAA expects that 50% more airline passengers will travel in 2013 than in 2001. FAA Administrator Blakey estimated at the 2003 Annual Forecast Conference that U.S. airlines would carry more than one billion passengers in 2014, compared to 628 million in 2002 (Blakey, 2003). This predicted movement of 1.1 billion passengers a year will require a safe and highly efficient Air Traffic Control (ATC) system. However, to meet the challenge of a growing air transport industry, the FAA must reduce delays and improve overall system efficiency. In congressional testimony, FAA Administrator Garvey stated "…the demand at LaGuardia exceeded supply at least 12 hrs a day. Delays of one to two hours or more were not unusual. Last September, delays were up 41 percent from the previous September and accounted for more than 20 percent of all delays nationwide" (Garvey, 2001a). In further testimony, the Administrator also announced, "…plans to improve operational efficiency at the eight airports with the highest delay rates" (Garvey, 2001b). These airports included Newark International (EWR), John F. Kennedy International (JFK), LaGuardia (LGA), and Philadelphia International (PHL) airports. The airspace of the New York Air Route Traffic Control Center (ARTCC; ZNY) and the New York Terminal Radar Approach Control (TRACON; N90) airspace is arguably the most complex and crowded in the world. This airspace is congested due to the combination of increasing air traffic volume and airspace complexity designed to accommodate the close proximity of JFK, LGA, and EWR airports. Changes in ATC procedures and airspace redesign have not kept pace with the concomitant increase in air traffic volume and complexity. Furthermore, there are no proposals to construct new runways at these airports for the next 10 to 15 years (National Airspace Redesign, 2002).

Figure 1 displays the current New York airspace configuration and adjacent facilities. As shown in Figure 1, there are several facilities adjacent to both ZNY and N90. Both ZNY and N90 are required to perform a considerable amount of coordination activity with each adjacent facility. ZNY must perform interfacility coordination with Cleveland (ZOB), Boston (ZBW), and Washington (ZDC) centers; ZNY Oceanic; and Philadelphia TRACON. N90 must coordinate with ZBW, ZDC, and ZNY Oceanic. Additionally, ZNY and N90 are also required to coordinate with each other. The number of adjacent facilities, coupled with the airspace complexity in general, requires a considerable amount of coordination activity to take place both within and between facilities.

In addition to high levels of communication and coordination, the complexity and current procedures of the New York airspace also have an impact on how Certified Professional Controllers (CPCs) are able to maneuver aircraft. The CPCs at ZNY and N90 are currently using procedures that have been modified only minimally over the last 30 years (National Airspace Redesign, 2002). Furthermore, the airspace complexity limits the CPCs' options for delay absorption and restricts close-in holding patterns. This restriction in turn creates uneven traffic feeds at the arrival fixes and results in gaps between aircraft on final approach.

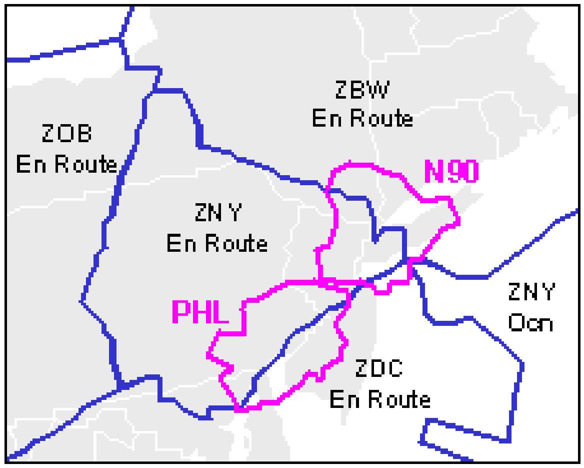


Figure 1. Current ZNY and N90 airspace boundaries and surrounding facilities.
Note: The blue lines indicate en route airspace and the fuchsia lines indicate terminal airspace.

Therefore, CPCs are unable to use runway capacity to its fullest extent. Additionally, the compact terminal airspace prevents the use of a more effective departure structure, and CPCs must implement static miles-in-trail metering. The relatively low terminal ceiling also requires aircraft to fly their holding patterns at lower altitudes, which can be inefficient and costly for the airlines.

1.1 Background

The New York Airspace Redesign Team has proposed airspace redesign and procedural changes to deal with the ever-increasing air traffic demand, delays, and airspace complexity. The Redesign Team, a group of ATC experts, have developed new procedures for the New York airspace with the goal of relieving some stress from the increasingly crowded airspace while improving performance and maintaining safety. The redesign team comprises representatives from the FAA Eastern Region Air Traffic Division Airspace Branch (AEA), along with Air Traffic Planning and Procedures (ATP-400), Terminal Business Service (ATB-300), Air Traffic Airspace Management (ATA-200), Research and Requirements Development Directorate (ARQ), and the National Air Traffic Controllers Association (NATCA) for ZNY and N90. Collectively, they have developed the New York Integrated Control Complex (NYICC) Concept of Operations (FAA, 2002b). The NYICC concept encompasses new procedures that may improve NAS efficiency within their domain. In part, the FAA authors state that the NYICC will improve NAS efficiency by reducing airspace complexity and unifying airspace to create more efficient arrival and departure routes. They expect that the NYICC concept would provide benefits in terms of traffic flow, intersector coordination, separation standards and methods, traffic sequencing and spacing, and holding.

If the FAA applied the NYICC concept, it would change the ZNY and N90 airspace by uniting these two facilities into a single, centralized facility. Figure 2 shows the proposed NYICC integrated airspace. In addition to airspace expansion that would include parts of ZBW and ZDC airspace, the collocation of ZNY and N90 operations has the potential advantage of reducing the amount of communication and coordination that must take place between the two entities via landline. For example, if adjacent N90 and ZNY sectors were close to one another on the control room floor, CPCs could coordinate and communicate face-to-face (FTF) regarding aircraft

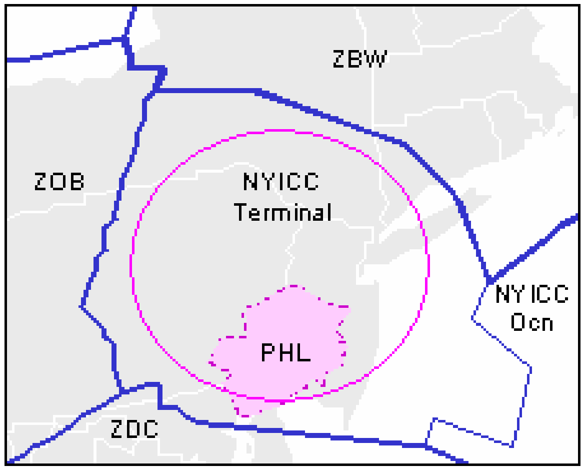


Figure 2. NYICC integrated airspace.

transitioning between the terminal and en route operations. The CPCs would also likely use gestures such as pointing and other nonverbal cues to facilitate communication. Furthermore, both the terminal and en route CPCs would be able to glance at the other's display to expand their awareness of the current and likely future situation.

Communication is essential to the operation of the air transport system (Orlady & Orlady, 1999). It is the medium for coordinating the overall system to maximize safety and efficiency and occurs at multiple levels (FAA with airlines, Air Traffic Control System Command Center (ATCSCC) with Air Traffic Management (ATM) at ARTCCs and TRACONs, ATM with controllers, intra- and intersector controller with controller, controller with pilot, pilot with pilot, and Air Traffic with Airway Facilities). There are many instances in which ineffective communication resulted in horrible consequences, such as the collision of two Boeing 747s in Tenerife, Spain (e.g., Hawkins, 1987), in which miscommunication between the tower and the aircraft resulted in 582 deaths. There are many other accidents in which miscommunication between the pilots was a causal factor, leading to the development of crew resource management programs (e.g., Jensen & Biegelski, 1989). There are also data demonstrating the effectiveness of successful coordination. After the implementation of collaborative decision making between the ATCSCC and the Airline Operations Centers, for example, more than two million minutes of delay were saved at 16 major airports during an approximate 6-month period in 1998-1999 by dynamically reallocating landing slots (Beatty, Corwin, & Wambsganss, 1999).

There are many factors that affect successful coordination. Obviously, sharing information in a timely manner is most important (e.g., MacDonald, 1998). Having a shared knowledge of the task and situation is also critical (e.g., Converse, Cannon-Bowers, & Salas, 1991). Of course, the method of communication is important. Some information is best conveyed orally, others visually (written or graphic), others nonverbally (e.g., gestures), or some combination of them (Orlady & Orlady, 1999). There is some research that indicates that FTF communication, where possible, is best for coordination (Warkentin, Sayeed, & Hightower, 1997). Rognin and

Blanquart (2000), in their analysis of the cognitive ergonomics of ATC, pointed out that collocated controllers “have the opportunity to observe each other, distributing and acquiring explicitly as well as implicitly information, through verbal messages, visual observation of other agents and of informational supports such as the radar screen, the strip progress board, the radio or the notepad” (p. 5). They also point out that collocation increases awareness about the current situation, about current actions, about the availability of the other, and about who is communicating. These past findings are supportive of the concept of collocating controllers whose sectors must interact with each other, regardless of whether they are en route or terminal sectors.

In addition to collocation, the NYICC concept also proposes to essentially extend the boundaries of the terminal airspace to effectively reduce the lateral separation standard within a larger portion of airspace. By extending the terminal airspace, there would be a reduction of the lateral separation from 5 nm to 3 nm in what were once en route sectors. The expanded terminal airspace has the potential to make the overall system more efficient by providing better arrival and departure routes. Better movement into and out of the terminal airspace may result in higher departure and landing rates, give CPCs the opportunity to make control actions sooner and more efficiently, and reduce the time an aircraft must stay aloft. The authors of the NYICC concept of operations (FAA, 2002b) foresee that the expanded terminal airspace would improve arrival and departure flows and reduce the number of transfer-of-control points. The potential efficiency gain resulting from the NYICC concept would be ideal if the result is a smoother flow of traffic and improved safety.

The proposed concepts of collocation and an expanded terminal airspace have the potential to improve NAS efficiency within the current ZNY and N90 airspace. Given the broad range of input and expertise utilized to generate the NYICC concept, there is reason to have a certain level of confidence in its success, despite the lack of empirical support. Human-in-the-loop simulation, controlled scientifically, can provide an empirical, yet realistic, test of the NYICC concept by identifying the potential benefits and risks of implementing these new procedures.

1.2 Purpose

The experiments described in this document provide an objective assessment concerning the use of alternative ATC procedures for arrivals and departures in N90 and ZNY. Engineering Research Psychologists (ERPs) from the William J. Hughes Technical Center (WJHTC) conducted a high fidelity, human-in-the-loop ATC simulation to scientifically compare current ATC procedures with two alternative procedures for directing air traffic to and from major New York airports. The experiments examined various aspects of CPC behavior and performance that occurred during each of three conditions. Based on the data obtained, we make recommendations regarding the alternative ATC procedures that we tested.

1.3 Scope

This document presents two separate experiments that used ZNY and N90 airspace. The first experiment examined arrival flow patterns, and the second experiment examined departure flow patterns. Each experiment independently assessed CPC workload, performance, communication behavior, and system efficiency in a realistic ATC human-in-the-loop simulation. The design and results of these experiments are limited to the ZNY and N90 airspace sectors, procedures, and experimental conditions that the experimenters actually used in each experiment.

2. METHOD – ARRIVAL EXPERIMENT

The following sections describe the experiment, that evaluated the NYICC concept on arrivals.

2.1 Participants

Eighteen male CPCs participated in the arrival simulation. Nine of the CPCs were from ZNY en route Sectors Broadway (74) and Milton (75); nine were from N90 Sectors Yardley/Penns (ARD) and EWR. The CPCs participated in three groups of six. Six of the participants, three from ZNY and three from N90, wore corrective lenses during the experiment. Table 1 shows the descriptive statistics (means and standard deviations [*SD*s]) for the biographic questionnaire data we collected from the nine en route and nine terminal CPC participants.

Table 1. Biographic Questionnaire Means and Standard Deviations for the En Route and Terminal CPC Participants

	En Route - ZNY	**Terminal – N90**
	Mean (*SD*)	**Mean (*SD*)**
What is your Age?	40.78 (3.53)	42.33 (3.67)
How long have you worked as a CPC (include both FAA and military experience)?	14.58 (2.69)	19.94 (4.69)
How long have you worked as a CPC for the FAA?	14.58 (2.69)	17.94 (2.97)
How long have you been a Certified Professional Controller (or Full Performance Level Controller)?	11.28 (2.83)	15.98 (2.75)
How long have you actively controlled traffic in the en route/terminal environment?	14.31 (2.80)	14.43 (4.82)
How many of the past 12 months have you actively controlled traffic?	12.00 (0.00)	12.00 (0.00)
Rate your current skill as a CPC.	9.22 (1.09)	6.44 (3.78)
Rate your current level of stress.	2.89 (2.20)	4.89 (3.62)
Rate your level of motivation to participate in this study.	7.89 (1.96)	5.89 (3.55)

Overall, the participants were highly experienced CPCs. The terminal participants were somewhat more experienced than the en route participants. En route participants rated their current skill level as being very high, whereas terminal participants rated their current skill level as being more moderate. The terminal participants' moderate self-rating (6.44 on a 10-point scale) of skill was in part due to two participants who rated their skill as extremely low (1 on a 10-point scale). These participants may have misinterpreted the rating scale because all of the controllers were highly experienced and certified to work complex traffic. The participants rated their current stress levels as low to moderate and their motivation to participate in the experiment as moderate to high.

2.2 Research Personnel

Three experimenters, a Principal Investigator (PI), Co-Principal Investigator (CPI), and one Research Assistant (RA) conducted the experiment. The PI was responsible for the overall management of the experiment including briefings, experimental procedure, data collection, and simulator preparation and operation. The CPI assisted in all of the PI activities and was prepared to assume all duties of the PI if necessary. The RA prepared experimental materials such as flight progress strips and assisted in the collection and analysis of data. Three Subject Matter Experts (SMEs) developed and prepared the scenarios for use in the simulation, in collaboration with representatives from the New York Airspace Redesign Team. These SMEs, along with one additional SME, also served as over-the-shoulder observers during the data collection phase of the experiment. CPCs from ZNY and N90 participated in the shakedown exercises to ensure that the scenarios were realistic. Hardware and software engineers prepared all equipment such as the ATC simulator, aircraft target generator, displays, and communications for use in the simulation. These engineers were on standby to assist if necessary during the experiment. Seventeen simulation pilots assisted with scenario shakedown and experimental runs.

2.3 Equipment

We conducted this experiment at the FAA WJHTC Research, Development, and Human Factors Laboratory (RDHFL) located at the Atlantic City International Airport, NJ. The RDHFL contained all equipment used.

2.3.1 Hardware

The CPC workstations and associated equipment were located in Experiment Rooms 1 and 2 of the RDHFL. The simulation pilot workstations were located in Experiment Room 3 and in the simulation pilot workstation room in the RDHFL.

2.3.1.1 Air Traffic Control Workstation Consoles

The experiment required two terminal and two en route radar workstation consoles. Each console contained a communication panel and flight strip bay. The ARD and the 75 consoles each had a keyboard and trackball for a Handoff position. Each en route console also contained a computer readout display (CRD). We did not include any of the controller decision support tools currently being field tested. Figure 3 presents a photograph of the workstation console configuration.

Figure 3. Photograph of terminal and en route workstation console configuration.

2.3.1.2 Simulation Pilot Workstations

We used 15 simulation pilot workstations in total to control the simulation aircraft. A single simulation pilot staffed each workstation. Each workstation consisted of a computer, keyboard, monitor, and communications equipment. Each simulation pilot had a plan view display of traffic and a list of assigned aircraft. For each assigned aircraft, the simulation pilots had information regarding the aircraft's current state and flight plan data. Each simulation pilot made entries at their workstation to effect clearances.

2.3.1.3 Communications

Communication panels and headsets were present at each participant and simulation pilot workstation. A participant controlling traffic in one sector could contact all other sectors involved in the simulation. The participants and the respective simulation pilots also had two-way voice communication via headsets.

2.3.1.4 Workload Assessment Keypad

The researchers placed Workload Assessment Keypads (WAKs) at each participant position, including handoff positions. We based the use of the WAK on prior research concerning the subjective measurement of CPC and pilot workload (Stein, 1985; Stein & Rosenberg, 1983). Researchers have used this or similar techniques to measure subjective workload in numerous experiments (e.g., Truitt, Durso, Crutchfield, Moertl, & Manning, 2000; Willems & Heiney, 2002; Willems & Truitt, 1999). Each of two laptop computers independently controlled three WAKs and provided data collection capability. Each WAK consists of a touch panel display with 10 numbered buttons. The WAK prompts the participant to press a button by both aural and visual signals. Experimenters can adjust the frequency of prompts (ratings) and the duration of time in which a participant has to make a response. During a prompt, the numbered buttons on the WAK illuminate and it emits a brief tone. The buttons will remain illuminated for the duration of the response period (20 s) or until the participant makes a response, whichever occurs first.

2.3.2 Software

The experimenters used the New Generation Target Generator Facility (TGF) and the Distributed Environment for Simulation, Rapid Engineering, and Experimentation (DESIREE) ATC simulator to present air traffic scenarios. Both were developed at the FAA WJHTC. The TGF supports simulations throughout the WJHTC. It uses flight plans to generate radar track and data block information on the controller and simulation pilot displays, and provides an interface for the simulation pilots to enter changes to the assigned heading, altitude, speed, or radio frequency. It also contains algorithms to control the aircraft maneuvers so that they represent how the actual aircraft would maneuver (e.g., realistic climb and turn rates). Finally, the TGF captures data about the aircraft trajectories, proximity to each other, and all the simulation pilot entries for later analyses.

DESIREE displays the radar tracks, history trails, data blocks, sector maps, navigation aids, lists, and so forth, on each controller radarscope. It also provides all the controller functionality, such as handoffs, point outs, entering data into the host computer, changing data block locations, and changing the display characteristics (e.g., range, brightness, and font size settings). DESIREE simulates the Standard Terminal Automation Replacement System (STARS) in the TRACON and the Display System Replacement (DSR) in the ARTCC. Finally, DESIREE has data collection capabilities, such as what data entries the controller made during a scenario. DESIREE was developed at the RDHFL to enable researchers to modify or add information displayed or functionality so they can evaluate different concepts or procedures. DESIREE and TGF are fully integrated systems.

2.4 Materials

2.4.1 Informed Consent Statement

Each participant read and signed an informed consent statement prior to the experiment (see Appendix A). The experimenters kept participant names separate from data to ensure confidentiality.

2.4.2 Biographic Questionnaire

Using the biographic questionnaire, the participants provided general information about themselves including gender, age, vision, and level of experience. The biographic questionnaire is shown in Appendix B.

2.4.3 NASA Task Load Index

Each participant completed the National Aeronautics and Space Administration (NASA) Task Load Index (TLX) (Hart & Staveland, 1988) ratings after each scenario to provide subjective ratings of workload, frustration, and performance. The NASA TLX rating form is shown in Appendix C.

2.4.4 Post-Scenario Questionnaire

The participants completed the Post-Scenario Questionnaire (PSQ) after each scenario to provide subjective ratings about their own performance, workload, and situation awareness (SA). The participants responded to each item of the PSQ by making a Likert scale rating from 1 to 10. The participants could also use the PSQ to provide open-ended responses to convey any information they considered relevant. The PSQ is shown in Appendix D.

2.4.5 Post-Experiment Questionnaire

The participants completed the Post-Experiment Questionnaire (PEQ) after completing all of the scenarios. Using the PEQ, the participants made Likert scale ratings (1-10) to provide their opinions regarding the general characteristics of the experiment. The participants could also use the PEQ to provide open-ended responses about any aspect of the experiment. The PEQ is shown in Appendix E.

2.4.6 Communication Score Sheet

An experimenter used the Communication Score Sheet (CSS) during each Collocated and Terminalized scenario to record verbal and nonverbal communication behavior between participants during each scenario. We developed the CSS for use in this particular experiment and based its content on prior research conducted by Peterson, Bailey, and Willems (2001). The CSS is shown in Appendix F.

2.4.7 ATC Observer Rating Form

The SMEs used either an en route or terminal Observer Rating Form (ORF), as appropriate, to record performance ratings for each participant. Both the terminal and en route ORFs are based on prior research conducted by Sollenberger, Stein, and Gromelski (1997). Appendix G shows the en route ORF and Appendix H shows the terminal ORF.

2.4.8 Standard Operating Procedures

With the exception of changes imposed by experimental conditions, the participants adhered to the actual Standard Operating Procedures (SOPs) that were currently in place at their respective facilities. This helped ensure that they used the same general procedures throughout the experiment and to bolster the validity of the experiment in general.

2.4.9 Scenarios

The SMEs, in collaboration with CPCs from N90 and ZNY, developed nine scenarios for use in the experiment. First, the SMEs created three scenarios based on actual recordings of ZNY and N90 air traffic that occurred on May 24, June 28, and July 19, 2001. The SMEs selected these three days because of the relatively high volume of traffic that occurred. The SMEs then increased the traffic levels for these three scenarios by 30%. We decided to increase the traffic levels from 100% to 130% so that we would be more likely to detect any effects due to the experimental procedures. Furthermore, the increased traffic load would approximate anticipated future traffic loads. The SMEs then created two more versions of each scenario by giving new callsigns to all aircraft. Therefore, the SMEs constructed three versions of each scenario (denoted as A, B, or C) resulting in nine total scenarios. All of the scenarios began with full traffic populating each sector and lasted for 50 min. The taskload due to air traffic remained relatively constant during each scenario; the number of aircraft did not decrease or "trail off" near the end of the scenarios.

2.5 Experimental Design

We essentially conducted two separate experiments within the arrival simulation. Because the CPCs were not qualified to control traffic in both terminal and en route sectors and because of the relatively small sample size, we treated the data collected within each airspace type separately. Even though the terminal and en route CPCs' actions likely affected one another, combining these data would have made the statistical analyses impossible. Therefore, for each airspace type (en route or terminal), the experiment used a 3 (Sector Position: 74, 75, 75H for en route, ARD, ARD H, EWR for terminal) X 3 (Condition: Normal, Collocated, Terminalized) repeated measures design. Each participant controlled traffic in each sector position and condition combination within their respective airspace type. While at each sector position, the participants experienced each of the three conditions (Normal, Collocated, Terminalized) once in succession. Table 2 displays the counterbalancing scheme that we used to control for any confounding effects that may have been due to the order in which participants experienced the various conditions or scenarios (see Appendix I).

Table 2. Arrival Experiment Condition and Scenario Counterbalancing

		En Route Sector			Terminal Sector				
Group	Simulation Run	74	75	75H	ARD	ARD H	EWR	Condition	Scenario
1	1	E1	E2	E3	T1	T2	T3	N	3A
"	2	"	"	"	"	"	"	C	1A
"	3	"	"	"	"	"	"	T	2A
"	4	E2	E3	E1	T2	T3	T1	C	3B
"	5	"	"	"	"	"	"	T	1B
"	6	"	"	"	"	"	"	N	2B
"	7	E3	E1	E2	T3	T1	T2	T	1C
"	8	"	"	"	"	"	"	N	2C
"	9	"	"	"	"	"	"	C	3C
2	10	E4	E5	E6	T4	T5	T6	T	1B
"	11	"	"	"	"	"	"	C	3B
"	12	"	"	"	"	"	"	N	2B
"	13	E6	E4	E5	T6	T4	T5	C	2C
"	14	"	"	"	"	"	"	N	1C
"	15	"	"	"	"	"	"	T	3C
"	16	E5	E6	E4	T5	T6	T4	N	3A
"	17	"	"	"	"	"	"	T	2A
"	18	"	"	"	"	"	"	C	1A
3	19	E7	E8	E9	T7	T8	T9	N	1C
"	20	"	"	"	"	"	"	C	2C
"	21	"	"	"	"	"	"	T	3C
"	22	E9	E7	E8	T9	T7	T8	T	2A
"	23	"	"	"	"	"	"	N	3A
"	24	"	"	"	"	"	"	C	1A
"	25	E8	E9	E7	T8	T9	T7	C	2B
"	26	"	"	"	"	"	"	N	1B
"	27	"	"	"	"	"	"	T	3B

(*Note*. E1 = Participant 1 for en route; T1 = Participant 1 for terminal; Condition N = Normal, Condition C = Collocated, Condition T = Terminalized)

2.5.1 Independent Variables

Both Sector Position (74, 75, 75H for en route and EWR, ARD, ARD H for terminal) and Condition (Normal, Collocated, Terminalized) were within-subject variables. During the Normal condition, the participants controlled traffic as they normally would in the field, and a wall physically separated the terminal and en route sectors. During the Collocated condition, the experimenters removed the wall from between the terminal and en route sectors making FTF communication between the sectors possible. During the Terminalized condition, the wall was not present (i.e., the Collocated condition) and we reduced the lateral separation standard for en route Sector 74 (adjacent to terminal Sector ARD) from 5 nm to 3 nm. We also reduced the radius of the J-ring from 5 nm to 3 nm.

2.5.2 Dependent Variables

For each condition, the experimenters and SMEs obtained measures of communication, performance, and efficiency. The participants provided subjective measures of performance, SA, and workload.

2.5.2.1 System Performance Measures

We collected numerous system performance measures for each sector of airspace. We designed these measures to provide information regarding the relative efficiency and safety of each experimental condition. The measures included number of altitude, heading, and airspeed commands issued; holding frequency and duration (time in seconds); throughput (i.e., aircraft count); total number of handoffs; time (in seconds) and distance flown (in nm) for all aircraft; number of flights handled (i.e., handoff taken and given for an aircraft); and number of aircraft landed. We also collected a number of error measures. The error measures included frequency and duration (time in seconds) of standard and user-defined conflicts (i.e., less than 3 nm lateral separation standard).

2.5.2.2 ATC Observer Rating Form

The SMEs used the terminal and en route ORFs to collect over-the-shoulder ratings for each participant occupying a radar or handoff position. Using the ORF, the SMEs provided an assessment of each participant's performance including their ability to maintain a safe and efficient traffic flow, maintain attention and SA, prioritize actions, provide control information, demonstrate technical knowledge, and to communicate. Each SME only made ORF ratings for a single sector.

2.5.2.3 Communication

We automatically recorded push-to-talk (PTT) landline communications by using a custom software application. We recorded the number of times each participant transmitted a message via the landline and the duration (seconds) of each transmission. Using the CSS form, an experimenter observed and recorded the frequency and general content of FTF verbal communications between en route Sector 74 and both ARD positions (radar side [R-side] and Handoff). The experimenter also recorded nonverbal gestures such as pointing to a display. In addition to the content and form of the communication, the experimenter recorded if a participant initiated the communication or was responding to a communication initiated by another participant.

We based all of the verbal communication-type descriptions on research conducted by Peterson et al., (2001). An Approval included communications about intersector control/approval requests (e.g., "Get me control for descent on that aircraft" or "APREQ N1234 climbing to FL330"). A Handoff included communications relating to the transfer of radar identification of a particular aircraft (e.g., "Handoff N1234" or "Did you handoff N1234?"). We coded a communication as a Point Out when it related to the transfer of radar identification of a particular aircraft when radio communications will be retained (e.g., "Point out N1234 to 22"). We coded communications about a traffic situation involving a specific aircraft including conflict, spacing, other protected air space or terrain, and the resolution of that situation (e.g., "Are you watching that aircraft") as Traffic. We coded communications about altitude not in relation to traffic (e.g., "N1234 is requesting flight level 220") as Altitude. We coded communications as Route when they regarded headings and/or amendments to route, not in relation to traffic situations (e.g., "N1234 is on a 330 heading" or "Next Sector, 27, wants N1234 over WEVER"). We coded communications about speed not in relation to traffic situations (e.g., "These three aircraft are slowed to 250 knots") as a Speed communication. The communications that we coded as Frequency were those about an aircraft's radio communications transfer or frequency assignment (e.g., "Have you switched N1234 yet?" or "Tell them to switch to 123.45"). We coded a communication as Flow Message if they were about traffic flow restrictions not referring to a specific aircraft (e.g., "The next sector is requesting 25 miles-in-trail"). The flight progress strip (FPS) communications included all communications about FPSs (e.g., "Where is that strip?"). Equipment communications included any communications about any ATC hardware (e.g., "The radar is out of service"). Finally, we coded communications as Aircraft Identification (A/C ID) when they involved identifying a specific aircraft (e.g., "Who was that who called?" or "That was N1234 who called").

We collected these measures using the CSS form during the Collocated and Terminalized conditions only because the participants were able to engage in FTF communication during these conditions. Furthermore, there was no a priori reason to hypothesize that communication behavior would change within terminal sectors or within en route sectors when we examined communication behavior during the Normal condition. However, there was reason to expect that landline communication between en route Sector 74 and terminal Sector ARD would decrease during the Collocated and Terminalized conditions in comparison to the Normal condition. We expected the participants to compensate for the hypothesized decrease in landline communication by using more FTF verbal and nonverbal communication during these conditions.

2.5.2.4 Workload

We used the WAK to obtain measures of subjective workload. The WAK prompted participants for a subjective workload rating every 5 min by emitting a brief, high-pitched chirp and illuminating the WAK buttons. Each participant had 20 s to respond to the prompt by pressing 1 of the 10 numbered buttons to indicate their current level of workload (1 = *low workload* to 10 = *high workload*). We recorded all workload ratings made by each participant. If a participant failed to respond within the 20-s time limit, then we considered the associated data points for that workload rating as missing data. We also administered a modified version of the NASA TLX after each scenario to obtain the participants' ratings of different dimensions of workload, frustration, and performance.

2.6 Procedure

2.6.1 General Schedule of Events

Each participant was involved in the arrival portion of the experiment for 2 days. Table 3 shows the daily schedule of events.

Table 3. Daily Schedule Of Events

First Day		Second Day	
Time	**Event**	**Time**	**Event**
8:30	Intro, Informed Consent, Bio	8:30	Equipment Familiarization
9:30	Equipment Familiarization	8:50	Break
10:00	Break	9:00	Simulation Run 5
10:15	Simulation Run 1	10:00	Break
11:15	Break	10:15	Simulation Run 6
11:30	Simulation Run 2	11:15	Break
12:30	Lunch	11:30	Equipment Familiarization
1:30	Simulation Run 3	11:50	Lunch
2:30	Break	1:00	Simulation Run 7
2:45	Equipment Familiarization	2:00	Break
3:05	Break	2:15	Simulation Run 8
3:15	Simulation Run 4	3:15	Break
4:15	Break	3:30	Simulation Run 9
4:30	Daily Out Briefing	4:30	Out Briefing, Questionnaires

2.6.2 Introductory Briefing

After being welcomed, the participants read and sign an informed consent statement. An experimenter and witness also signed the informed consent statement. We reassured the participants that they would not experience any adversities and that they were free to withdraw from the experiment at any time. The participants then completed the biographic questionnaire to provide information about their experience as a CPC. The PI then gave a short briefing to present the schedule of events and to explain the general procedures of the experiment. The briefing covered topics such as the length of scenarios and some of the dependent measures we would collect (e.g., WAK, observer ratings, PSQ, PEQ). After the experimenter's briefing, the SMEs briefed the participants on the hardware and SOPs that they would be using during the simulations. The SMEs also described the idiosyncrasies that existed between the simulated environment and an actual ATC position. The participants then familiarized themselves with the equipment and practiced adjusting display preferences before data collection began. During this familiarization process, we presented a sample air traffic scenario to the participants.

2.6.3 Data Collection Procedure

Throughout the entire data collection procedure, participant names were not associated with any data. Data collection began after we completed the introductory briefing. First, the participants received general instructions about the experimental condition that they would experience (Normal, Collocated, Terminalized). For the Normal condition, an experimenter instructed the participants that they were to control traffic as they normally would in the field. For the Collocated condition, an experimenter told the participants that the en route and terminal operations would be physically located next to one another and that they may use FTF communication if they wish. For the Terminalized condition, an experimenter instructed the participants that en route and terminal operations would be collocated and separation minimums would decrease from 5 nm to 3 nm lateral separation for en route Sector 74. Next, the participants received instructions about the WAK device before each scenario to refresh their memory and to increase the likelihood that they would use the same rating criteria over the course of the experiment. Appendix J contains the instruction set.

The appropriate air traffic scenario began after the participants had received all instructions and the experimenters had answered any questions. The participants experienced the different air traffic scenarios according to the counterbalancing scheme shown in Table 2. Each scenario began with a position relief briefing given to the participants by the SME assigned to their sector. Once the position relief briefings were complete, the participants took full control of the scenario. All scenarios were 50 min in duration.

During each air traffic scenario, the experimenters and SMEs collected the dependent measures described in Section 2.5.2. An experimenter used automated methods to collect audio, video, PTT, and WAK data. The SMEs used the ORFs to record performance ratings of participants working at each sector. The participants provided subjective ratings of workload using the WAK at 5 min intervals during each scenario. As soon as the scenario ended, the participants completed the NASA TLX and PSQ, in that order. The participants then took a break for at least 15 min before the next air traffic scenario began.

Using the counterbalancing scheme shown in Table 2, an experimenter assigned each participant to a new position within their domain (terminal or en route) after every three simulation runs. Before participants began controlling traffic at the new position, they took as much time as needed to familiarize themselves with the equipment at that position. For each of three groups of participants, we repeated this general procedure over two consecutive days for nine simulation runs per group.

The participants completed the PEQ after their group completed nine simulation runs. The participants, experimenters, and SMEs then gathered in the RDHFL briefing room for a caucus. During the caucus, the PI debriefed the participants regarding the nature of the experiment and answered any questions. The PI and SMEs solicited comments from the participants regarding the Collocated and Terminalized conditions. The participants also provided comments about the experiment in general. After the debriefing, we thanked the participants and released them from the facility.

3. RESULTS – ARRIVAL EXPERIMENT

In the arrival experiment, data were analyzed using the appropriate repeated measures analysis of variance (ANOVA) procedure for each dataset (see Appendix I for information on repeated measure designs and our overall approach to analyzing these data). We analyzed data collected from the terminal and en route sectors separately. If we found a significant interaction or main effect, then we computed the Tukey's Honestly Significant Difference (HSD) test to identify the differences. However, because the HSD can be less powerful than a simple effects ANOVA and we have a relatively small sample size, in some cases we obtained a significant main effect or interaction, but the post hoc HSD was not significant. In these cases, we assume that only the highest and lowest means are significantly different. Furthermore, by using the more conservative HSD post hoc test, we simplify the analyses and avoid capitalizing on chance by increasing the probability of a Type I error (false rejection of the null hypothesis).

We only report statistically significant effects when they are relevant to the question at hand. For example, a significant interaction that results from differences between Sector Position 74 in the Normal condition and Sector Position 75 in the Collocated condition may not be of practical interest in terms of the questions we are asking in this experiment. Therefore, we would not report a significant interaction of this type.

All statistically significant results reported in this document are significant at $p \leq .05$ unless stated otherwise. Because of our relatively low statistical power to detect significant effects, we also report marginal effects ($p \leq .10$) to help explain trends in the data. These marginal effects are important for the interpretation of the data even though they don't provide the same strength of support as the effects we arbitrarily call "significant." Whereas there is a 5% probability that a significant result was due to chance when using a criteria of $p \leq .05$, the probability of finding a significant result by chance increases to 10%when using a criteria of $p \leq .10$. When the data are presented graphically, the columns represent the mean value for the condition, position, or both. The error bars extending from the top of the columns represent the variability of the measure across the participants.

3.1 Simulation Realism

We first examine the participants' opinions of the overall realism of the simulation. Providing a high-fidelity simulation is important for at least two reasons. First, it is easier to generalize the results to the real world when the simulated conditions closely approximate actual conditions. This is the notion of external validity as discussed by Campbell and Stanley (1963). Second, the participants may be more motivated to take part in a simulation that is realistic and has face validity. The participants provided their opinion of the simulation realism by responding to six items on the PEQ using a 10-point scale (see Table 4). Overall, the participants rated the realism of the simulation as being moderate to high. The lower ratings were associated with the hardware and software, which are different from what they normally use. The ratings for the airspace and traffic scenarios were high, which is the most important aspect of simulation fidelity. They also indicated that the WAK rating procedure caused little interference with their ATC performance.

Table 4. Means and Standard Deviations for the PEQ Items on Simulation Realism

		Terminal – N90	**En Route - ZNY**
Item No.	**PEQ Item**	**Mean (*SD*)**	**Mean (*SD*)**
5	Rate the realism of the overall simulation experience compared to actual ATC operations.	5.89 (2.26)	5.44 (1.94)
6	Rate the realism of the simulation hardware compared to actual equipment.	5.33 (2.92)	6.44 (2.07)
7	Rate the realism of the simulation software compared to actual functionality.	5.56 (2.01)	5.78 (1.85)
8	Rate the realism of the simulation traffic scenarios compared to actual NAS traffic.	7.00 (1.50)	6.44 (1.94)
9	Rate the realism of the simulation airspace compared to actual NAS airspace.	8.67 (1.50	8.78 (1.30)
10	To what extent did the WAK online workload rating technique interfere with your ATC performance?	2.00 (2.65)	2.22 (1.39)

3.2 System Performance Measures

In 1999, the Free Flight Program Office (FFPO) implemented a plan to evaluate the effects of FFPO tools on NAS performance (FAA, 1999). The tools include the User Request Evaluation Tool (URET) and the Traffic Management Advisor (TMA). URET is an en route data side (D-side) tool that provides conflict alerts and allows trial planning of conflict solutions to ensure they are free of other conflicts. It can also be used to test whether pilot route or altitude requests are conflict free. TMA is designed to assist en route controllers in selecting time schedules to TRACON meter fixes to optimize arrival throughput. They were initially deployed to a limited number of facilities before deciding on wider deployment.

The evaluation plan for FFPO tools identified numerous metrics in the performance categories of safety, user access, delay/efficiency, predictability, flexibility, and system productivity. The measures included changes in arrival rates, flying time and distance, percentage of time at or near desired altitude, fuel used, monitor alert thresholds, operational errors, and so on, when comparing performance before and after the tools were deployed to field sites. The authors of the evolution plan recognized that many of these measures would be difficult to obtain and to interpret; NAS performance can be affected by many variables (e.g., weather, airport configuration), the value of different aspects vary over time (e.g., predictability versus flexibility), some actions may appear to be negative but were really positive (e.g., a reroute may result in longer flight times and distances, which can be measured, but it resulted in a smoother ride, which is difficult to determine), and the impact of the tools may only occur during peak periods. Consequently, they collected data over extended periods of time (typically 1 year before deployment as a baseline and continuously after that) and attempted to adjust for or consider the multiple factors that could affect their data.

Beginning in December 2000, the FFPO began issuing a performance report every 6 months to evaluate the impact of the tools and provide the information to all stakeholders. When possible, they conducted statistical analyses to determine if changes were significant, but in many cases, only descriptive data were presented. In some cases, the data were not appropriate for statistical analyses, and in other cases, the statistical analysis would not be meaningful, but the descriptive data had practical importance. For example, operational errors occur so rarely that there is no appropriate statistical model for evaluating changes attributable to the tools. For the latter case, changes in average distance flown across an entire center may be so small that it is statistically not significant but very important to the stakeholders. After URET was able to make direct lateral flight plan amendments into the Host computer system, for example, they found a reduction in average flying distance of 1 mile per flight during the peak hours of the day, which was not statistically significant. However, that difference translated into a fuel savings to the airlines of $1,875,000 per month (FAA, 2001).

After the introduction of TMA, the initial data indicated an increase of one to two aircraft arrivals per hour and an average decrease in delay of 1.63 min at peak times (FAA, 2001). Subsequent analyses (FAA, 2002c) have found even greater benefits, especially when confounding factors were statistically controlled. At one airport, there was initially no difference before and after TMA was introduced at the respective ARTCC. After a repeater display was installed in the TRACON, there was an increase of 1.4 arrivals for instrument approaches and 0.7 arrivals for visual approaches. At another airport, the arrival rate increased after TMA by about one per hour for instrument approaches and approximately three per hour for visual approaches in one airport configuration. At yet another airport, the mean peak actual arrival rates increased from about 62 per hour to about 66 per hour with the introduction of TMA. The ARTCC and TRACON personnel attributed the results to “increased situational awareness, better coordination between the facilities, proper front-loading, and shorter and more tactical miles-in-trail restrictions.”

These findings demonstrate the effects of changes to the NAS on aircraft flights, ATC operations, and airports. They also demonstrate that the changes are relatively small and are not necessarily statistically significant, but that they have a cumulative impact on an overburdened system that is of high practical significance. Some readers may consider many of the differences that we report here relatively small. However, in the ATC system, these small differences will sum over time and may result in a significant operational outcome. Although the results obtained in the previous examples of URET and TMA could have occurred theoretically by chance, a consistent positive result, whether statistically significant or not, usually indicates that the effects are systematic. Given the reasons just stated, we primarily report summary descriptive statistics for the system performance metrics and only use statistical inference when appropriate.

We collected all of the system data during twenty-seven 50 min simulation scenarios. During each simulation run, we collected system performance measures from files created by the TGF software. The system performance measures are objective measures of what the aircraft actually did during each simulation run. We collected the system performance measures for each sector position including the Ghost sector. The Ghost sector is important to consider in some of the following analyses because it represented activity that would have taken place in the adjacent ZDC and ZOB airspace. We based all system performance measures on what we refer to as “frequency-based” calculations. That is, we calculated the system measures based on the sector radio frequency an aircraft was using to receive and transmit radio communications. Therefore, for purposes of the following analyses, an aircraft was transferred from one sector to another when the participant instructed an aircraft to contact the next sector and the simulation pilot actually switched frequencies. We describe the system performance measures and report the summary descriptive statistics in the following sections.

3.2.1 Number of Completed Flights

We calculated the number of completed flights by counting the number of aircraft that descended below an altitude of 1200 ft msl. As an aircraft descended to 1200 ft msl, the participant would have already given control of the aircraft to a tower controller. Therefore, we treat these aircraft as having landed for the purposes of this data analysis.[1]

---

[1] We lost approximately 10 min of data for the third simulation run completed by Group Three (Terminalized condition). A computer hardware malfunction caused the loss of data. For the number of completed flights variable, we predicted the missing values for this run by determining how many completed flights occurred during each of the first ten 3.5-min segments of the simulation run. Then, using a time series analysis, we determined the values for the next four 3.5 min time intervals. We used the resultant value to replace the biased mean in the abbreviated simulation run. We accounted for all other missing system data by using grand mean replacement.

Overall, the participants were able to complete more flights on average in the Terminalized condition compared to both the Normal and Collocated conditions (see Figure 4). In the 50 min scenarios, the participants were able to complete 2.13 more flights on average in the Terminalized condition compared to the Normal condition. The participants completed 2.58 more flights on average in the Terminalized condition compared to the Collocated condition. Although the statistical analysis was not significant, these numbers may be operationally significant. Compared to the Normal condition, the Terminalized condition resulted in a 7.1% increase in the number of flights completed over the same period.

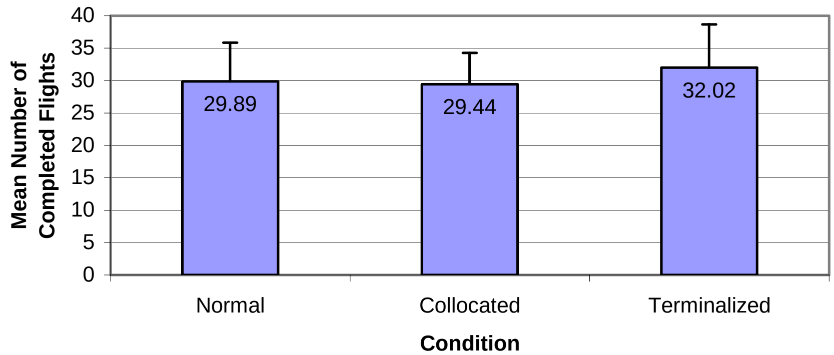


Figure 4. Mean number of completed flights by Condition.

3.2.2 Number of Aircraft Handled

We calculated the number of aircraft handled by counting, for each sector, the number of aircraft that were on the frequency when a scenario began plus the number of handoffs that participants accepted during a scenario. We considered the number of aircraft handled as a measure of efficiency in that procedures that are more efficient should allow participants to handle more aircraft. When we examined the mean number of aircraft handled by Condition, it appeared that participants were able to handle the most aircraft in the Collocated and Terminalized conditions (see Figure 5). Compared to the Normal condition, participants handled 1.56 more aircraft on average during the Collocated condition. While in the Terminalized condition, participants handled 1.77 more aircraft on average compared to the Normal condition.

To better understand how this overall benefit of the Collocated and Terminalized conditions occurred, we examined system performance in each of the simulated sectors. As Figure 6 shows, participants in the ARD, 74, and 75 sectors demonstrated an incremental increase in the number of aircraft handled. In these sectors, the participants handled the fewest number of aircraft on average in the Normal condition and the greatest number of aircraft in the Terminalized condition. The participants handled an intermediate number of aircraft in the Collocated condition. The en route participants in Sectors 74 and 75 realized the greatest benefit of the Terminalized condition with an average increase of 3.79 and 4.30, respectively, in the mean number of aircraft handled. Conversely, terminal participants who worked the EWR sector handled the greatest number of aircraft during the Collocated condition, followed by the Normal and then Terminalized conditions. The terminal participants' benefit during the Collocated condition was somewhat smaller than that realized by the en route participants, though. There was virtually no change in the number of aircraft handled by the Ghost sector, indicating that the overall effects were due to changes in the controlled sector positions and not due to changes in the uncontrolled Ghost position.

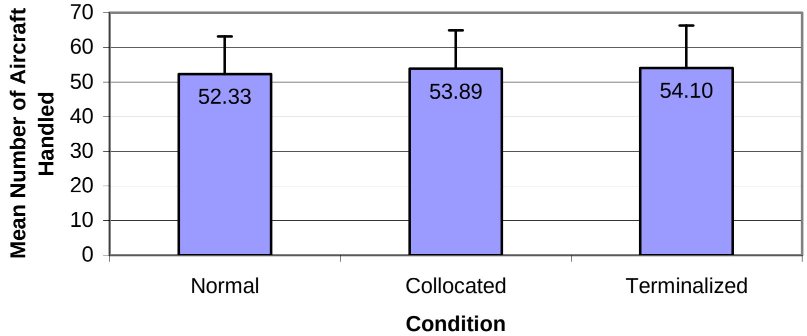
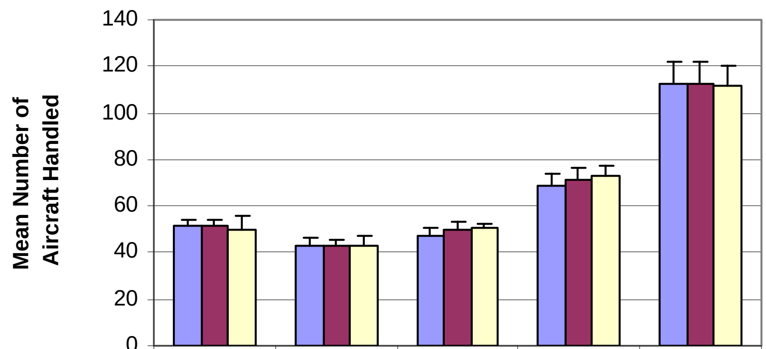


Figure 5. Mean number of aircraft handled collapsed across all Sector Positions (EWR, ARD, 74, and 75) by Condition.



	EWR	ARD	74	75	GHOST
Normal	51.11	42.67	47.00	68.56	112.67
Collocated	51.89	43.11	49.67	70.89	112.67
Terminalized	49.42	43.34	50.79	72.86	111.47

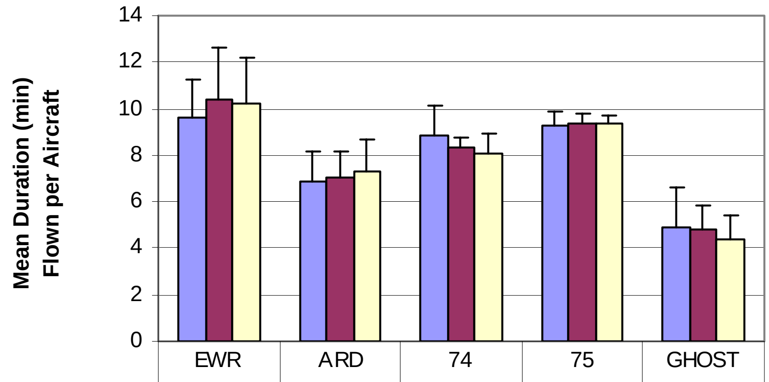
Sector Position

Figure 6. Mean number of aircraft handled by Sector Position and Condition.

3.2.3 Duration of Aircraft Handled

In addition to the number of aircraft handled, we also examined the duration that participants handled each aircraft by measuring the time each aircraft was on each sector's radio frequency.

Figure 7 shows the mean duration (min) each aircraft was on the radio frequency by Sector Position and Condition. These data show that the condition affected each sector differently.



	EWR	ARD	74	75	GHOST
Normal	9.63	6.89	8.83	9.28	4.89
Collocated	10.35	7.04	8.29	9.32	4.78
Terminalized	10.22	7.31	8.08	9.35	4.39

Figure 7. Mean duration (min) on the radio frequency per aircraft by Sector Position and Condition.

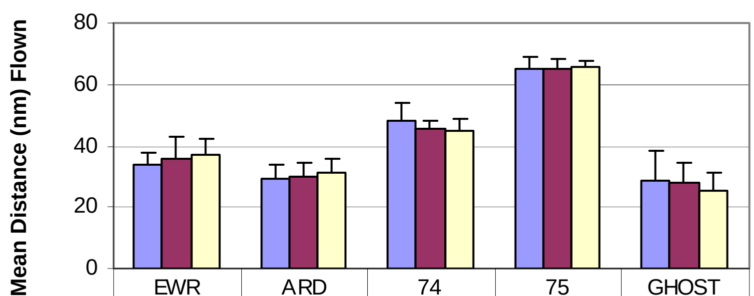
The data trend from the EWR sector indicates an increase in the duration handled per aircraft for the Collocated and Terminalized conditions. For the EWR sector, the participants had each aircraft on frequency about 43.2 s longer on average during the Collocated condition than during the Normal condition. The aircraft were on frequency about 35.4 s longer during the Terminalized condition. The increased duration for the Terminalized condition occurred despite the fact that EWR handled relatively fewer aircraft in this condition. When participants worked the ARD sector, they also kept aircraft on their frequency slightly longer during the Collocated and Terminalized conditions. In the Terminalized condition, the participants kept aircraft on frequency about 25.2 s longer than during the Normal condition. During the Collocated condition, the participants kept aircraft on frequency only about 9 s longer compared to the Normal condition.

In contrast to the terminal sectors, en route Sector 74 reduced the time they were in radio contact with aircraft during the Normal condition by 45 s during the Terminalized condition, and by 32.4 s in the Collocated condition. The differences between conditions were very small for Sector 75. Like Sector 74, the duration aircraft flew on frequency in the Ghost sector decreases from the Normal to the Collocated to the Terminalized condition. Aircraft flew on the Ghost sector frequency for an average of 30 s less in the Terminalized condition compared to the Normal condition.

Overall, en route participants at Sector 75 were taking handoffs sooner during the Terminalized condition. However, once in Sector 75, aircraft were spending the same amount of time in that sector regardless of condition. It also appears from the data that terminal participants may have been taking handoffs from Sector 74 sooner during the Collocated and Terminalized conditions compared to the Normal condition. Therefore, the reduction in mean duration flown on frequency for en route sectors may have translated into higher duration on frequency for the terminal sectors. Both ARD and EWR appear to have experienced a slight increase in the mean duration of time each aircraft was on their respective frequencies. An alternate explanation of these data would be that participants at EWR and ARD were experiencing slightly higher taskload during the Collocated and Terminalized conditions because they were handling more aircraft overall. Therefore, the increase in duration flown on frequency may have been due to the participants handing off arrival aircraft to the tower later than usual.

3.2.4 Distance Flown

The distance-flown variable measured how far the aircraft flew during each simulation run. We calculated the mean distance flown per aircraft to examine the overall efficiency in each experimental condition. We calculated this measure for each sector and condition by dividing the total distance flown by the number of aircraft handled. Figure 8 shows the descriptive statistics. We measured an increase in the distance flown for the EWR and ARD sectors in the Collocated and Terminalized conditions, with the greatest distance flown occurring in the Terminalized condition. Conversely, the opposite effect obtained for the en route Sector 74 and the Ghost sector. Sector 75 showed a slight increase in the distance flown during the Terminalized condition. Like the measure of duration handled, there are similar possible explanations for these data. The distance flown may have decreased in the Ghost sector due to less holding in the Collocated and Terminalized conditions or because the participants at Sector 75 were taking handoffs sooner. The distance flown in Sector 74 may have been due to more efficient flight paths or the participants at ARD may have been accepting handoffs from Sector 74 sooner. The increase in distance flown in the ARD and EWR sectors may have been due to the participants at these sectors taking handoffs sooner or because less efficient routes were needed to accommodate an increase in the number of aircraft handled. The EWR participants may have also been handing the aircraft to the tower later in the Collocated and Terminalized conditions due to an increase in taskload created by an increase in the number of aircraft handled.



	EWR	ARD	74	75	GHOST
Normal	33.99	29.09	48.16	64.76	28.89
Collocated	35.94	29.85	45.63	64.89	28.15
Terminalized	36.77	30.96	45.02	65.58	25.10

Sector Position

Figure 8. Mean distance (nm) flown per aircraft by Condition and Sector.

3.2.5 Number of Holds

Throughout the experiment, the participants had to put some aircraft into holding patterns during each simulation run. The holding was necessary because of the high traffic load that we presented to the participants. According to the NYICC Concept of Operations (FAA, 2002b), the authors hypothesized that collocation and expanded terminal airspace would result in more efficient traffic flows and less holding. Figure 9 shows the mean number of holds summed across Sector Position by Condition. To obtain these numbers, we calculated the mean number of holds that occurred for each Sector Position and Condition combination. Then, for each Condition, we summed the mean number of holds that occurred. As the data show, the most holds occurred during the Collocated condition, whereas the fewest number of holds occurred in the Terminalized condition. Apparently, it was the combination of collocation and an expanded terminal airspace that resulted in the greatest benefit in terms of number of holds.

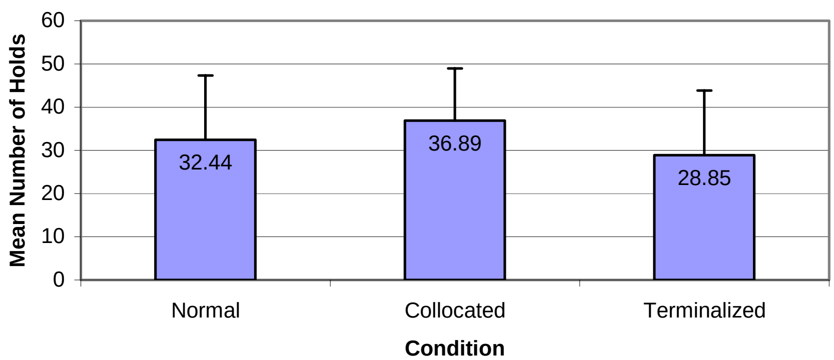
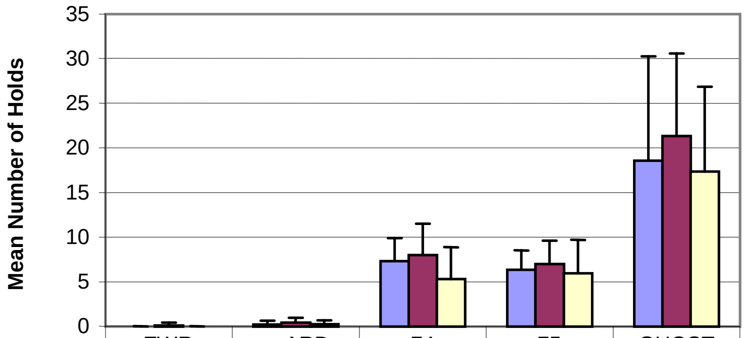


Figure 9. Mean number of holds summed across Sector Position by Condition.

When we examine the mean number of holds by Sector Position and Condition (shown in Figure 10), we see the same basic pattern in the individual sectors that occurred in the overall data. There was very little holding that occurred in the EWR and ARD sectors, however, Sectors 74, 75, and Ghost each had the most holding activity during the Collocated condition and the least holding in the Terminalized condition. There was a lot of variability in these data, but the overall data trend remained consistent. These data also show that most of the holding occurred in the Ghost sector that represented ZDC and ZOB. Therefore, the participants held numerous aircraft before they entered ZNY.



	EWR	ARD	74	75	GHOST
Normal	0.000	0.222	7.333	6.333	18.556
Collocated	0.111	0.444	8.000	7.000	21.333
Terminalized	0.000	0.256	5.321	5.936	17.342

Sector Position

Figure 10. Mean number of holds by Sector Position and Condition.

3.2.6 Duration of Holds

In addition to the number of holds, we also calculated the duration of holds. Figure 11 shows the mean cumulative duration of holds by Condition. We calculated this variable by summing the mean duration of holds that occurred over all simulation runs in each Condition.

Cumulatively, the participants had aircraft in the holding pattern longest during the Normal condition, even though there were more aircraft in holds during the Collocated condition. The participants had aircraft in the holding pattern for the least amount of cumulative time in the Terminalized condition.

We also calculated the mean duration of hold per aircraft by dividing the mean cumulative duration by the mean number of aircraft held in each Condition. Like the mean cumulative duration of holds, the mean duration of holds per aircraft also supported the hypothesis that holding would be more efficient in the Collocated and Terminalized conditions (see Figure 12).

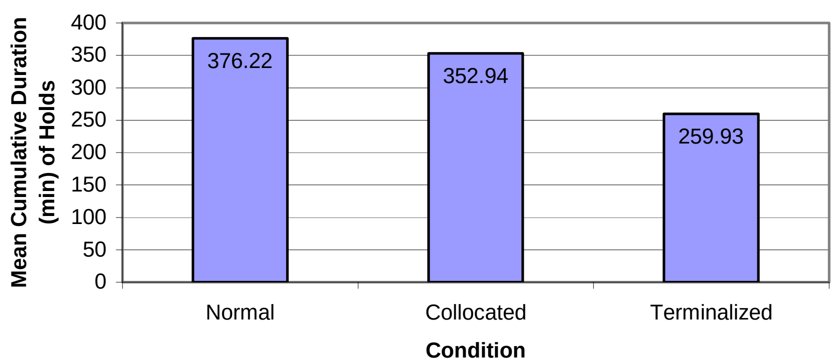


Figure 11. Mean cumulative duration (min) of holds by Condition.

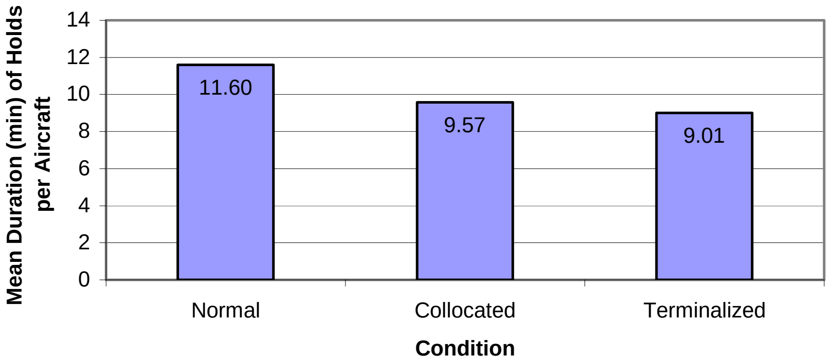


Figure 12. Mean duration (min) of holds per aircraft by Condition.

On average, holding aircraft spent the longest amount of time in the pattern during the Normal condition. Compared to the Normal condition, aircraft spent about 2 min less in the Collocated condition and 2.6 min less in the holding pattern during the Terminalized condition.

The same data trend that exists for the summary data also exists when we examine the mean duration of holds per aircraft by Sector Position and Condition (see Figure 13). For the en route Sectors 74 and 75 and the Ghost sector, the mean duration of holds tended to be longest in the Normal condition and shortest in the Terminalized condition. For the terminal Sectors EWR and ARD, the mean duration of holds tended to be longer in the Collocated condition. However, very little holding occurred in either of the terminal sectors.

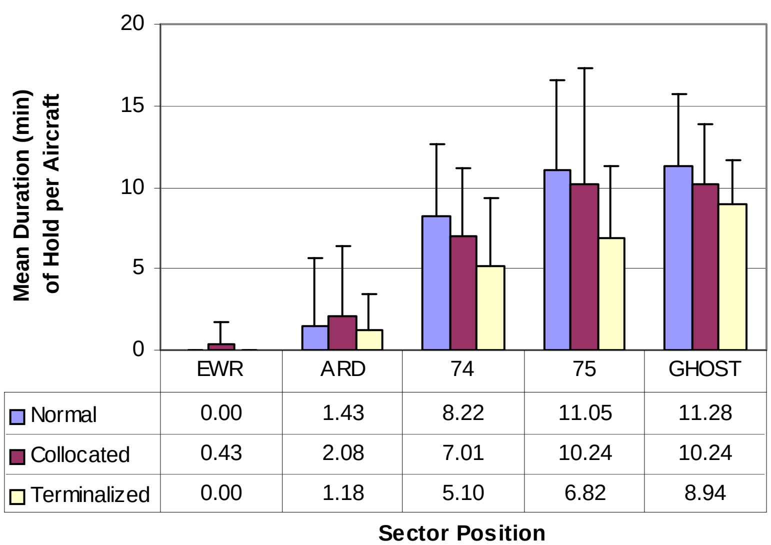
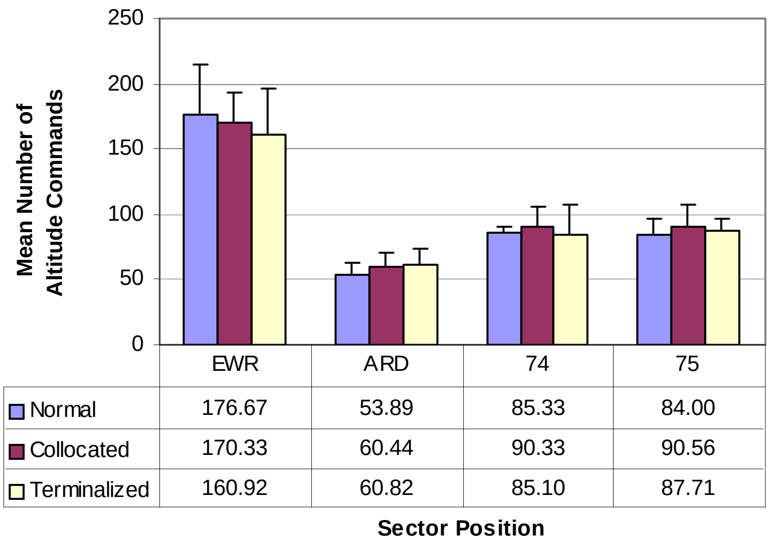


Figure 13. Mean duration (min) of holds per aircraft by Sector Position and Condition.

3.2.7 Number of Altitude Commands

The participants gave about 400 altitude commands in each condition. The system performance data for the mean number of altitude commands showed different patterns for the different types of sectors (see Figure 14). We chose to show the data in this way to convey the absolute number of altitude commands that the participants were making. Although the number of aircraft that the participants handled during each condition confounds these data, we can see that taskload increased in some sectors under certain conditions but decreased in other sectors. The en route Sectors 74 and 75 tended to give the most altitude commands during the Collocated condition. For the terminal Sector ARD, the participants issued more altitude commands in the Terminalized and Collocated conditions than in the Normal condition. The EWR sector, on the other hand, had the fewest number of altitude commands in the Terminalized condition and the most in the Normal condition. The terminal sectors may have experienced a trade off in that the participants made altitude adjustments more frequently in the ARD sector during the Collocated and Terminalized conditions, whereas the participants in the EWR sector made fewer during these conditions. This suggests that the Terminalized condition, and the Collocated condition to a lesser extent, allowed participants to set up aircraft for the arrival sooner, compared to the Normal condition.



	EWR	ARD	74	75
Normal	176.67	53.89	85.33	84.00
Collocated	170.33	60.44	90.33	90.56
Terminalized	160.92	60.82	85.10	87.71

Figure 14. Mean cumulative number of altitude commands by Sector Position and Condition.

The data shown in Figure 14 provide information regarding the participants' overall taskload that may be attributed to altitude commands. However, it is difficult to determine from these data if the Collocated or Terminalized conditions provided any benefits in terms of efficiency because they do not account for the number of aircraft that the participants handled. Figure 15 shows the mean number of altitude commands per aircraft by Sector Position and Condition.

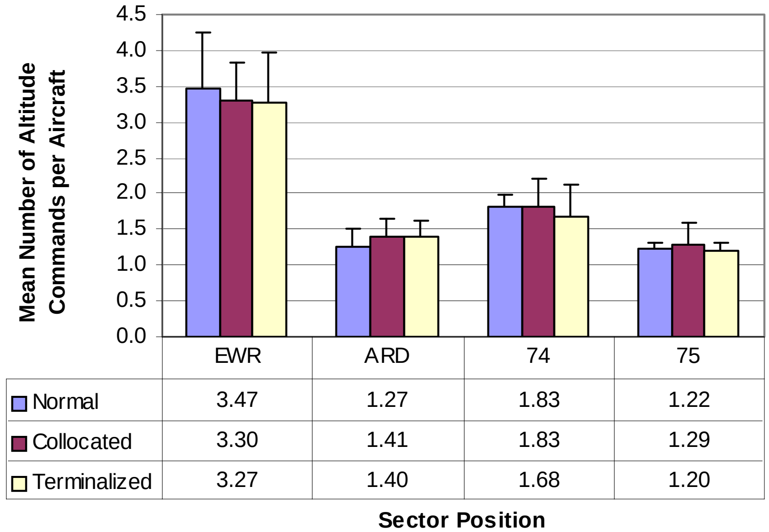
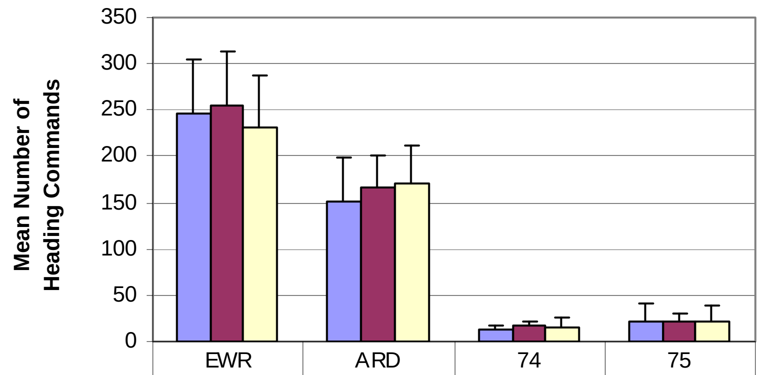


Figure 15. Mean number of altitude commands per aircraft by Sector Position and Condition.

The mean number of altitude commands did change slightly across conditions within each sector position, but there was too much variability in the data to determine if these differences were significant. The data trend does resemble the trend that obtained for the cumulative number of altitude commands. The participants at EWR seem to be giving fewer altitude commands per aircraft in the Collocated and Terminalized conditions; however, the participants at ARD seem to be giving more in these conditions.

3.2.8 Number of Heading Commands

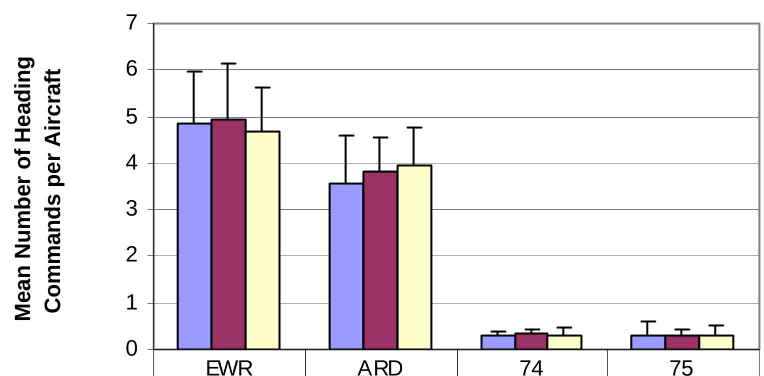
On average, the participants gave about 443 heading commands during each scenario. As with the altitude commands, we chose to examine the total number of altitude commands that the participants made to assess how taskload might have changed across sector positions and conditions due to the number of heading commands. Although the number of aircraft that the participants handled during each condition confounds these data, we can see that taskload may have increased in some sectors under certain conditions but seemed to decrease in other sectors (see Figure 16). The participants gave most of the heading commands in the EWR sector. While working at the EWR sector, the participants gave fewer heading commands in the Terminalized condition. Conversely, the participants at the ARD sector tended to give more heading commands in the Terminalized condition. Again, a tradeoff in taskload between the two terminal sectors may have occurred. The participants who worked at en route Sectors 74 and 75 gave relatively few heading commands in all conditions.



	EWR	ARD	74	75
Normal	247.11	151.78	12.89	21.00
Collocated	255.44	165.56	16.22	21.67
Terminalized	230.14	171.21	14.74	22.53

Figure 16. Mean cumulative number of heading commands by Sector Position and Condition.

We also calculated the mean number of heading commands per aircraft (see Figure 17). The experimental conditions did not seem to affect Sectors 74 and 75. There was a relatively large amount of variability in the data for the EWR and ARD sector positions. Overall, neither the Collocated nor the Terminalized conditions seemed to affect the mean number of heading commands given per aircraft. However, the data trend does resemble the trend that we obtained



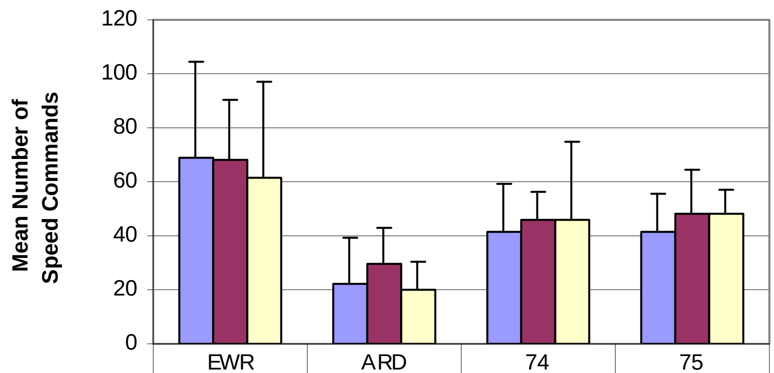
	EWR	ARD	74	75
Normal	4.84	3.56	0.28	0.31
Collocated	4.94	3.83	0.33	0.31
Terminalized	4.66	3.94	0.29	0.31

Figure 17. Mean number of heading commands per aircraft by Sector Position and Condition.

for the cumulative number of heading commands. The participants at EWR seemed to be giving fewer heading commands per aircraft in the Terminalized condition, whereas the participants at ARD seemed to be giving more.

3.2.9 Number of Speed Commands

The participants gave about 180 speed commands during each scenario. We examined the total number of speed commands that the participants made to assess taskload due to the number of speed commands. Again, we can see that taskload may have increased in some sectors under certain conditions, but it seemed to decrease in other sectors (see Figure 18). The participants working at the EWR sector tended to give more speed commands but were able to reduce this number slightly in the Terminalized condition. When the participants worked the ARD sector, they tended to give more speed commands during the Collocated condition. Of all sector positions, the fewest speed commands occurred at the ARD sector position. The en route participants tended to give a relatively moderate amount of speed commands compared to the terminal sectors, and they gave fewer speed commands in the Normal condition. For the en route sectors, this result suggests that there may have been a slight increase in taskload due to the number of speed commands in the Collocated and Terminalized conditions.

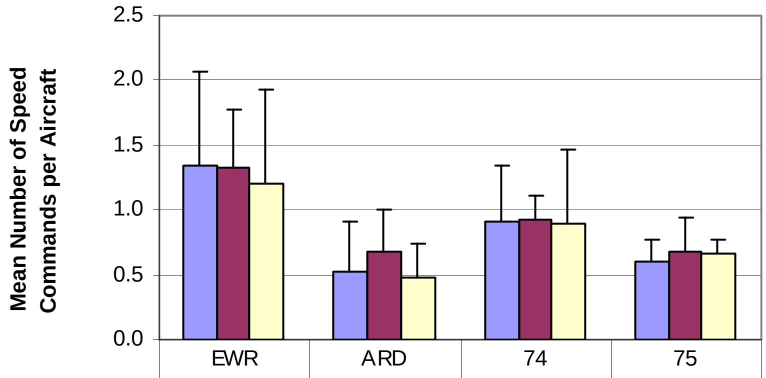


	EWR	ARD	74	75
Normal	68.67	22.56	41.67	41.22
Collocated	68.33	29.33	45.78	47.89
Terminalized	61.46	19.88	45.94	48.31

Sector Position

Figure 18. Mean cumulative number of speed commands by Sector Position and Condition.

To account for the number of aircraft that the participants handled, we also examined the mean number of speed commands per aircraft (see Figure 19). The number of speed commands per aircraft showed the same general pattern as the cumulative number of speed commands.



	EWR	ARD	74	75
Normal	1.35	0.53	0.91	0.59
Collocated	1.32	0.68	0.92	0.69
Terminalized	1.21	0.47	0.90	0.66

Figure 19. Mean number of speed commands per aircraft by Sector Position and Condition.

Overall, the Collocated condition appeared to increase taskload somewhat compared to the Normal and Terminalized conditions because the participants gave more altitude, heading, and speed commands during this condition. Taskload may have increased slightly during the Collocated condition, but the participants were also handling more aircraft than in the Normal condition. However, the participants were also handling more aircraft in the Terminalized condition compared to the Normal condition, but there was not a corresponding increase in the overall number of commands. The participants gave most of the altitude, heading, and speed commands at the EWR and ARD sectors. Compared to the Normal condition, there were only small changes in the number of commands per aircraft in the Collocated and Terminalized conditions. There was also much more variability between the participants than between the conditions.

3.3 Safety Measures

To assess the potential air traffic risk in each condition, we examined how often pairs of aircraft lost standard separation and how often operational errors occurred. We also examined the frequency of wake turbulence violations in each condition. Wake turbulence violations are of particular interest for the en route participants. En route controllers use a 5 nm lateral separation standard and are not generally concerned with the effects of wake turbulence. However, in the Terminalized condition of this experiment, the en route participants in Sector 74 used a 3 nm lateral separation standard and, therefore, had to consider the effects of wake turbulence.

3.3.1 Loss of Separation

Researchers determined the loss of separation using the separation standards as defined in the FAA Order 7110.65N (FAA, 2002d) with the following exceptions noted. For the en route Sectors, 74 and 75, if the aircraft were above an altitude of 29,000 ft msl (i.e., flight level (FL) 290), we determined loss of separation when any two or more aircraft came within 5 nm laterally and 2,000 ft vertically of each other. If the aircraft were below FL 290, we determined loss of separation when aircraft came within 5 nm laterally and 1,000 ft vertically of each other. For the terminal Sectors EWR and ARD, we defined loss of separation as any instance where two or more aircraft came within 3 nm laterally and 900 ft vertically of each other. We also applied these terminal separation standards to the en route Sector 74 during the Terminalized condition. We used these rules only to provide an approximate index of risk. In calculating the loss of separation, we did not account for aircraft that were on diverging headings or aircraft that may have been using visual separation procedures.

We performed separate analyses for the en route and terminal loss of separation data. For each data set, we conducted a 2 (Sector) X 3 (Condition) repeated measures ANOVA. For the en route sectors, there was a significant Sector X Condition interaction, $F(2, 16) = 5.54$. For Sector 74, there was a significant reduction in the number of loss of separation occurrences in the Terminalized condition compared to the Normal condition, HSD(16) = 7.46. Figure 20 shows the en route loss of separation data.

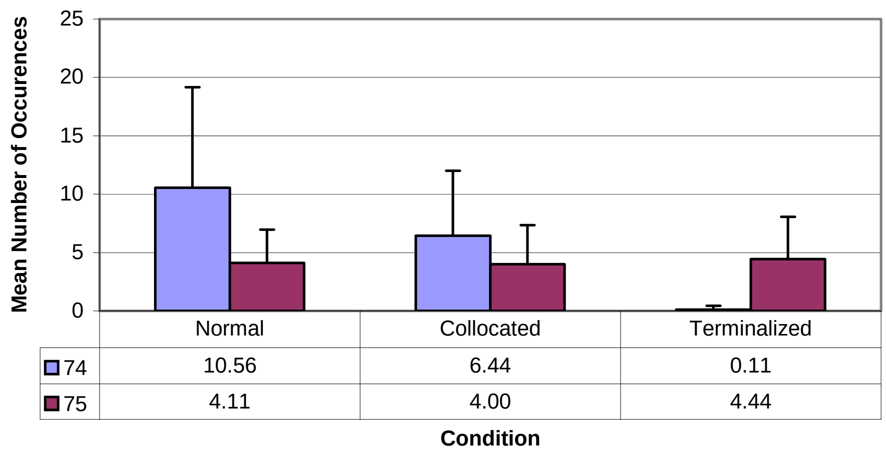


Figure 20. Mean number of occurrences of en route loss of separation by Sector and Condition.

For the terminal sectors, there was a significant main effect of Sector, $F(1, 8) = 36.13$. There was significantly more loss of separation instances in the EWR sector compared to the ARD sector, HSD(8) = 3.50. We expected the greater number of loss of separation incidents in the EWR sector because of the greater need for aircraft maneuvering and spacing on final approach. Furthermore, we did not account for reduced spacing requirements that may apply within 10 nm of the landing runway. Figure 21 shows the terminal loss of separation data.

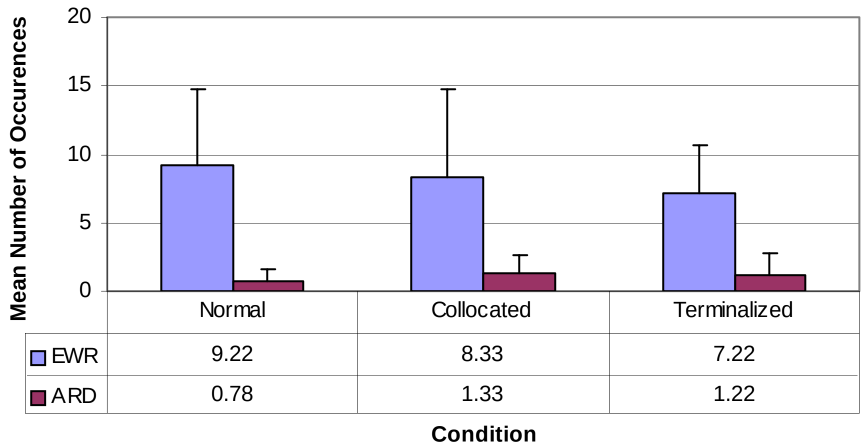


Figure 21. Mean number of occurrences of terminal loss of separation by Sector and Condition.

We also examined the number of en route operational errors. For a loss of separation to be classified as an operational error, an aircraft pair in which one or both aircraft were not level had to come within 4.8 nm and 1900 ft of each other if they were above FL 290, or within 4.8 nm and 900 ft of each other if they were below FL 290. If both of the aircraft were level, they had to come within 5 nm and 1700 ft of each other if they were above FL 290, or within 5 nm and 700 ft of each other if they were below FL 290. In calculating operational errors, we only used these simple criteria. Our analysis of operational errors was not as sophisticated as the investigation that would occur in the field. However, we did review each instance that we classified as an en route operational error by examining multiple data sources. We referred to SME notes and the audio/video recordings to establish the likelihood that an operational error occurred. These data sources provided information that allowed us to ensure that an extraneous factor such as a malfunction of the TGF, DESIREE, or a simulation pilot error did not cause any of the operational errors. We only classified an operational error as such once we ruled out possible extraneous factors. We were unable to statistically analyze the en route operational error data because there was not any variability (i.e., no operational errors) in Sector 74 during any of the scenarios using the Normal condition. Figure 22 shows the en route operational error data.

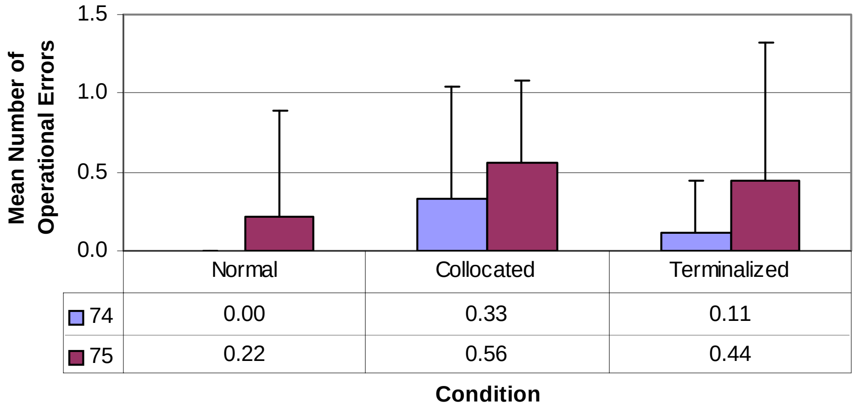


Figure 22. Mean number of operational errors by Sector and Condition.

Overall, we classified very few instances as operational errors. Although it appears that more operational errors occurred in the Collocated and Terminalized conditions, the absolute number is too small to draw any inferences. In addition, any effects associated with the experimental conditions may be a result of unfamiliarity with them rather than the design.

3.3.2 Wake Turbulence

We calculated wake turbulence violations based on FAA Order 7110.65N (FAA, 2002d) with some exceptions. We only examined aircraft pairs that were above 3000 ft msl. We assumed that aircraft below that altitude were the responsibility of the tower controller. In addition, for us to categorize an aircraft pair as being in violation of the wake turbulence separation standard, the headings of the aircraft pair had to differ by less than 90 degrees because of the software used to reduce the data. Therefore, our criteria for categorizing a wake turbulence violation was more liberal than the actual FAA standards that allow separation to be maintained given diverging courses of 15 degrees or more. That is, we detected more violations than would occur under operational conditions.

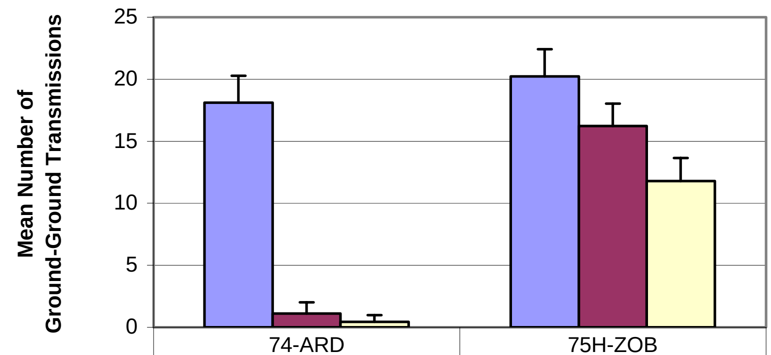
There were significantly more wake turbulence violations in the EWR sector (Mean = 18.04, SD = 2.14) compared to the ARD sector (Mean = 2.19, SD = 0.99), $F(1, 8) = 42.09$. Wake turbulence violations occurred equally often in all conditions for both the en route and terminal sectors.

3.4 Communication Measures

For communications behavior, we examined the PTT and the FTF communication data separately for both the en route and terminal participants. For the PTT data, we examined the mean number and duration of ground-ground and ground-air transmissions within each domain. The ground-ground transmissions were those landline transmissions that went between either two participants or between a participant and simulation pilot at one of the Ghost sectors (ZOB or ZDC). The ground-air transmissions were those transmissions that went from a participant to a simulation pilot that was flying an aircraft inside the ZNY or N90 airspace.

3.4.1 En Route Push-to-Talk Communications

The en route ground-ground transmissions for the PTT data include landline transmissions made from Sector Position 74 to ARD and transmissions made from Sector Position 75H to ZOB. We analyzed the number of en route ground-ground transmissions using a 2 (Sector Position: 74, 75H) X 3 (Condition: Normal, Collocated, Terminalized) repeated measures ANOVA. There was a significant main effect of Condition, $F(2, 16) = 13.09$. The Condition X Sector Position interaction was marginal. The post hoc test for Condition indicated that there were significantly more landline communications during the Normal condition than either the Collocated or Terminalized conditions, HSD(16) = 9.87 (see Figure 23).



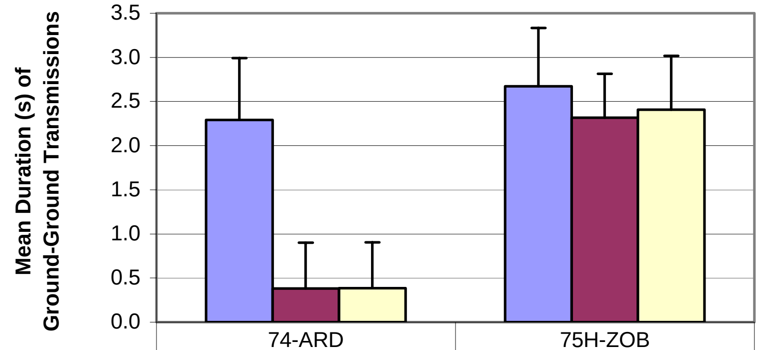
	74-ARD	75H-ZOB
Normal	18.11	20.22
Collocated	1.11	16.22
Terminalized	0.44	11.78

Sector Position

Figure 23. Mean number of en route ground-ground landline transmissions by Sector Position and Condition.

The marginal interaction indicates that the Collocated and Terminalized conditions affected Sector Position 74 more than Sector Position 75H. The finding that ground-ground transmissions decreased in the Collocated and Terminalized conditions for Sector Position 74 was not surprising because the wall that separated Sector Positions 74 and ARD was not present during these conditions. The coordination and communication that participants would normally conduct via landline could now be conducted in an FTF manner. The reduced number of transmissions between Sector Position 75H and ZOB most likely resulted from the decreased number of holds that participants performed during the Collocated and Terminalized conditions.

There was a significant Sector Position X Condition interaction, $F(2, 16) = 4.86$, for the mean duration of en route ground-ground transmissions (see Figure 24). This interaction suggests that the condition affected each en route sector position differently. The significant post hoc test, HSD(16) = 1.35, states that the participants at Sector Position 74 decreased the mean duration of their ground-ground transmissions more than did the participants at Sector Position 75H during the Collocated and Terminalized conditions. Furthermore, the participants at Sector Position 74 also significantly decreased the mean duration of their ground-ground transmission during the Collocated and Terminalized conditions compared to the Normal condition.



	74-ARD	75H-ZOB
Normal	2.29	2.67
Collocated	0.38	2.32
Terminalized	0.38	2.41

Sector Position

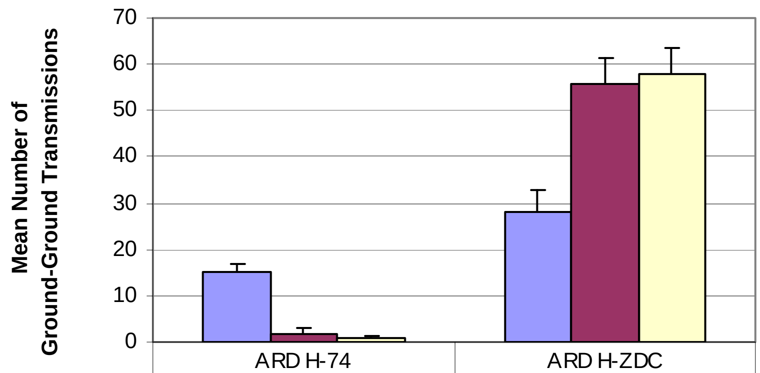
Figure 24. Mean duration (s) of en route ground-ground landline transmissions by Sector Position and Condition.

The ground-air transmissions for the en route PTT data includes radio transmissions made from Sector Positions 74 and 75 to a pilot. The analysis indicated a significant main effect of Sector Position, $F(1, 8) = 22.40$. The participants at Sector Position 75 (Mean = 278.81, $SD = 3.55$) made more transmissions to pilots than did the participants at Sector Position 74 (Mean = 230.19, $SD = 3.32$). We expected that Sector Position 75 would make more ground-air transmissions because they handled more aircraft compared to Sector Position 74.

There were no significant effects for the mean duration of en route ground-air transmissions. On average, each transmission made by a participant took about 3.5 s.

3.4.2 Terminal Push-to-Talk Communications

The ground-ground transmissions for the terminal PTT data include landline transmissions made from Sector Position ARD H to Sector Position 74 and to ZDC. We analyzed the terminal ground-ground transmissions using a 2 (Sector Position: ARD H-74, ARD H-ZDC) X 3 (Condition: Normal, Collocated, Terminalized) repeated measures ANOVA. The analysis of the mean number of terminal ground-ground transmissions did not detect any significant differences, but did return a marginal Sector Position X Condition interaction. The analysis did not find a significant effect here due to a large amount of variability in the data. The experimental conditions appeared to affect the participants' ground-ground communication behavior at the ARD sector Position. The participants reduced their communication with Sector Position 74 in the Collocated and Terminalized conditions, but the frequency of their communication with ZDC increased (see Figure 25).

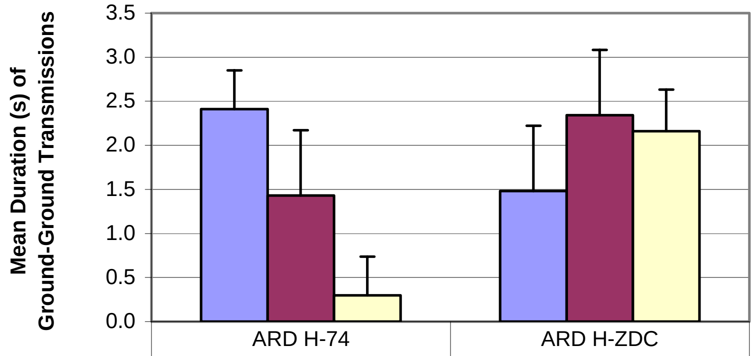


	ARD H-74	ARD H-ZDC
Normal	15.00	28.11
Collocated	1.89	55.67
Terminalized	0.67	58.11

Sector Position

Figure 25. Mean number of terminal ground-ground landline transmissions by Sector Position and Condition.

There was a significant Sector Position X Condition interaction for the mean duration of terminal ground-ground transmissions, $F(2, 16) = 5.69$ (see Figure 26). The participants reduced the mean duration of their communications between Sector Position ARD H and 74 in the Terminalized condition compared to the Normal condition, HSD(16) = 1.92. The duration of communications between Sector Position ARD H and ZDC appeared to increase in the Collocated and Terminalized conditions, but this difference was not statistically significant.



	ARD H-74	ARD H-ZDC
Normal	2.41	1.48
Collocated	1.43	2.34
Terminalized	0.30	2.16

Sector Position

Figure 26. Mean duration of terminal ground-ground landline transmissions by Sector Position and Condition.

The ground-air transmissions for the terminal PTT data includes radio transmissions made from Sector Positions EWR and ARD to a pilot. We analyzed the terminal air-ground transmissions using a 2 (Sector Position: EWR, ARD) X 3 (Condition: Normal, Collocated, Terminalized) repeated measures ANOVA. For the mean number of terminal ground-air transmissions, there was a significant main effect of Sector Position only, $F(1, 8) = 147.94$. The participants at the EWR sector position (Mean = 390.30, *SD* = 5.60) made significantly more ground-air radio transmissions than did the participants at the ARD sector position (Mean = 253.56, *SD* = 5.14). The mean number of ground-air transmissions was relatively stable across conditions. However, the participants at the EWR sector position may have realized a small, although not statistically significant, benefit during the Terminalized condition in that they made 24 fewer ground-air transmissions on average during this condition compared to the Normal condition and 34 fewer transmissions compared to the Collocated condition.

The analysis of the mean duration of the terminal ground-air transmissions did not find any significant effects. On average, the participants' ground-air transmissions took about 3 s each. The mean duration of these transmissions remained stable across Sector Positions and Conditions.

Overall, the independent variable of condition tended to have the greatest effect on the mean number and duration of ground-ground landline transmissions. This was true for both en route and terminal sector positions. The amount of landline activity between the en route and terminal Sector Positions 74 and ARD tended to decrease during the Collocated and Terminalized conditions when compared to the Normal condition. Furthermore, when landline communication did take place between Sector Positions 74 and ARD during the Collocated and Terminalized conditions, it tended to be of shorter duration on average compared to the Normal condition. The number of ground-ground transmissions and the duration of those transmissions also decreased between Sector Position 75H and ZOB during the Collocated and Terminalized conditions compared to the Normal condition. Conversely, the mean number and duration of ground-ground transmissions increased between Sector Position ARD and ZDC during the Collocated and Terminalized conditions. The Collocated and Terminalized conditions may have reduced the number and duration of intrafacility landline communications, but they may have also increased the number and duration of interfacility landline communications between N90 and the adjacent ZDC. In contrast to the ground-ground landline transmissions, the condition had very little or no effect on the number or duration of ground-air radio transmissions in the en route and terminal sector positions.

3.4.3 Face-to-Face Communications

The participants may have compensated for the reduction in the number of landline communications during the Collocated and Terminalized conditions, especially those communications between Sector Positions ARD and 74, by engaging in more FTF communication. FTF communication between the terminal and en route sectors was possible only in the Collocated and Terminalized conditions because experimenters removed the wall separating terminal and en route sector positions during these two conditions. In this section, we examine the communication data collected using the CSS.

The participants did take advantage of the opportunity for FTF communication. On average, there were 38.78 verbal communications in the Collocated condition and 40.56 verbal communications in the Terminalized condition. These FTF verbal communications more than made up for the decrease in landline communication, which fell from a mean of 33.11 in the Normal condition to 3.00 in the Collocated condition and 1.11 in the Terminalized condition. In addition to exchanging verbal information, they also used the opportunity to acquire information by looking (glancing) at one another's radar display. Table 5 shows the mean number of basic communication behaviors by Condition and Type of behavior. We recorded a Glance whenever a participant at Sector Position ARD looked over at Sector Position 74's radar display or whenever the participant at Sector Position 74 looked over at ARD's display. Verbal communications included any type of ATC-related FTF communication. Non-verbal communications included gestures such as pointing, giving a "thumbs up," or nodding one's head in acknowledgment. Communications that we coded as Other included ATC-related communications for which the experimenter was unable to understand the meaning. Non-ATC communications were those that went between Sector Position ARD and 74 but were not related to ATC whatsoever. The communications that the observer classified as Could Not Code were those that the observer could not hear or otherwise understand. Most of the FTF communications were verbal, followed closely by glances. The observer recorded relatively few behaviors in the remaining categories.

Table 5. Mean Number of Basic Communication Behaviors by Condition and Type

	Glance	Verbal	Non-verbal	Other	Non-ATC	Could not code
Collocated	26.44	38.78	3.22	2.11	2.11	0.67
Terminalized	31.56	40.56	3.67	3.89	1.67	0.56

A more detailed examination of the types of verbal communications that took place is shown in Table 6. Overall, the most frequent type of FTF communications was those regarding Traffic or Flow Messages. There were very few communications regarding Approvals, Point Outs, or FPS. No FTF communication took place that we would have coded as Frequency or A/C ID, so we do not show these categories.

Table 6. Mean Number of Verbal Communication Behaviors by Condition and Type

	Approval	Handoff	Point Out	Traffic	Altitude	Route	Speed	Flow Msg	FPS	Equipment
Collocated	0.00	1.78	0.11	14.00	3.89	2.78	1.11	13.33	0.00	1.78
Terminalized	0.22	0.78	0.00	9.67	5.78	1.78	4.56	16.67	0.33	0.78

We can further examine the participants' FTF communication behavior by understanding who was directing the communication to whom. Table 7 shows the mean number of basic communication behaviors by Condition, Sector Position, and Type. We coded whether communications were going from Sector Positions ARD or ARD H to 74 or from Sector Position 74 to ARD/ARD H. We did not distinguish if communications were going from Sector Position 74 to ARD or ARD H because either participant at the ARD radar or handoff positions were just as likely to receive the communication and we did not want to make any assumptions about who actually received the communication. The data in Table 7 indicate that the participants at all three sector positions took advantage of their collocation in both the Collocated and Terminalized conditions. Not only did they exchange information verbally, but they also used the collocation to acquire information from each other's radar display. The participants at Sector Positions ARD and ARD H in particular tended to glance at Sector Position 74's radar display to acquire information. We expected this result to some extent because the aircraft are arriving into EWR via Sector 74 and then ARD in that order. The data in Table 7 also indicate that the participants exchanged verbal communications in both directions.

Table 7. Mean Number of Basic Communication Behaviors by Condition, Sector Position, and Type

		Glance	Verbal	Non-verbal	Other	Non-ATC	Could not code
Collocated	A/P←74	5.89	16.78	0.33	0.89	1.22	0.44
	A/P→74	9.33	9.22	1.22	0.56	0.22	0.22
	A/PH→74	11.22	12.78	1.67	0.67	0.67	0.00
Terminalized	A/P←74	6.78	18.78	0.89	1.56	0.67	0.22
	A/P→74	11.56	8.11	1.44	1.11	0.56	0.33
	A/PH→74	13.22	13.67	1.33	1.22	0.44	0.00

Overall, the participants, both terminal and en route, took advantage of their collocated situation during the Collocated and Terminalized conditions by engaging in FTF communication. They used this FTF communication instead of making landline transmissions to one another. The participants also took advantage of their collocated situation by glancing at each other's radar display and gaining information from that display. This was especially true for the terminal participants at Sector Positions ARD and ARD H who were concerned with the arrival flow of traffic into their sector from Sector 74.

3.5 ATC Observer Rating Form

The SMEs rated the participants' performance using the ORF. They made their ratings using an 8-point scale where a rating of 1 indicated the *least effective performance* and a rating of 8 indicated the *most effective performance*. There are different items on the terminal ORF and the en route ORF. We analyzed the SME ratings from the ORF for the en route and terminal participants separately.

3.5.1 En Route Observer Rating Form

Overall, the SME ratings on all of the ORF items indicated high levels of performance. To determine if there were any differential effects, we analyzed each item of the en route ORF using a 3 (Sector Position) X 3 (Condition) repeated measures ANOVA. We report the statistically significant effects below.

There was a significant main effect of Condition for the participants' rated ability to take actions in the appropriate order of importance, $F(2, 14) = 6.48$. The post hoc test was not significant, but the data trend suggests that the mean ORF ratings were highest during the Terminalized condition (see Figure 27). However, the ratings for this item were high under all conditions.

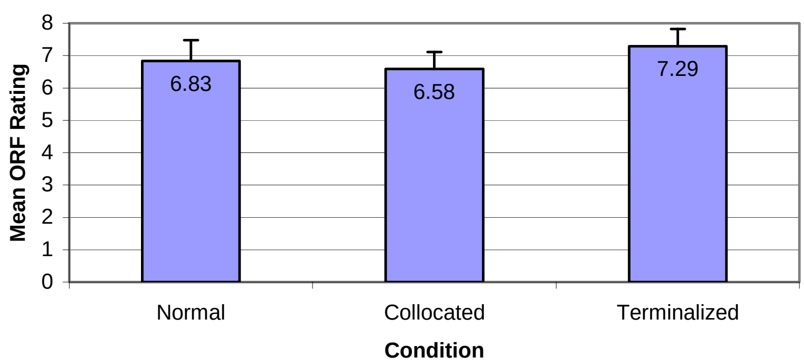


Figure 27. Mean en route ORF ratings for taking actions in appropriate order of importance by Condition.

There was a significant main effect of Condition for the participants' rated ability to preplan for control actions, $F(2, 14) = 7.40$. The post hoc test was not significant. The participants' rated ability to preplan control actions was good to very good in all conditions. However, the data trend suggests that the mean ORF ratings were highest for the Terminalized condition (see Figure 28).

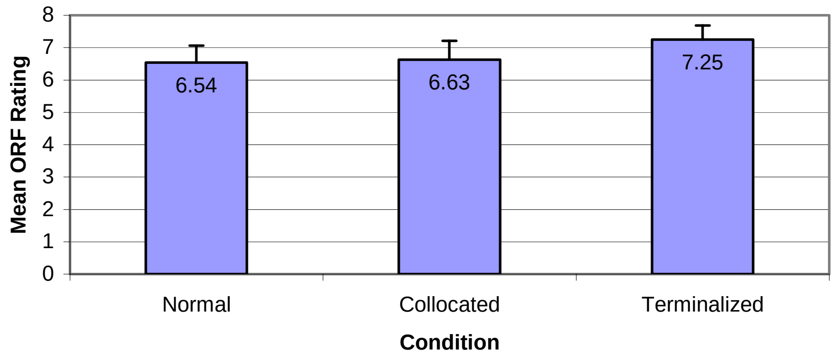


Figure 28. Mean en route ORF ratings for preplanning for control actions by Condition.

We also found a significant effect of Condition for the participants' rated FPS marking activity while performing other actions, $F(2, 14) = 7.86$. The post hoc test was not significant. The ratings were good to very good in all conditions, but they were highest in the Terminalized condition (see Figure 29).

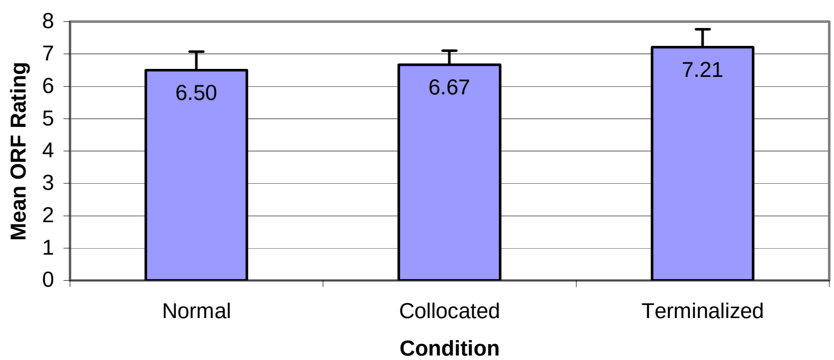


Figure 29. Mean en route ORF ratings for marking flight progress strips while performing other actions by Condition.

There was a significant main effect of Condition for the participants' rated overall ability to prioritize, $F(2, 12) = 13.05$ (see Figure 30). The post hoc test was not significant but a marginal effect ($p < .10$) suggests that the mean ORF ratings were higher for the Terminalized condition compared to both the Normal and Collocated conditions. Therefore, the data indicate that participants may have been better able to prioritize their activities during the Terminalized condition. Alternatively, participants may have had more activities to prioritize during the Terminalized condition due to the increased number of aircraft handled, and the SME raters noticed this.

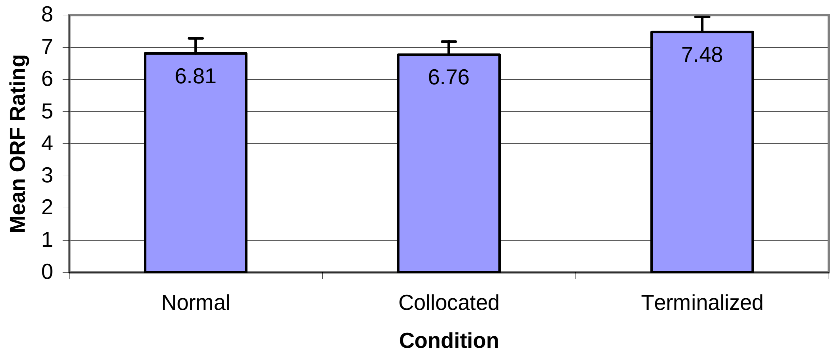


Figure 30. Mean en route ORF ratings for overall prioritizing by Condition.

We found a significant main effect of Condition and Sector Position for the participants' rated ability to provide control information, $F(2, 12) = 4.56$ and $F(2, 12) = 5.01$, respectively. Neither post hoc test was significant. However, the data trend suggests that the mean ratings were higher in the Terminalized condition and when the participants were working Sector 74 (see Figure 31). The participants' rated ability to provide control information was very good to most effective in all conditions and positions.

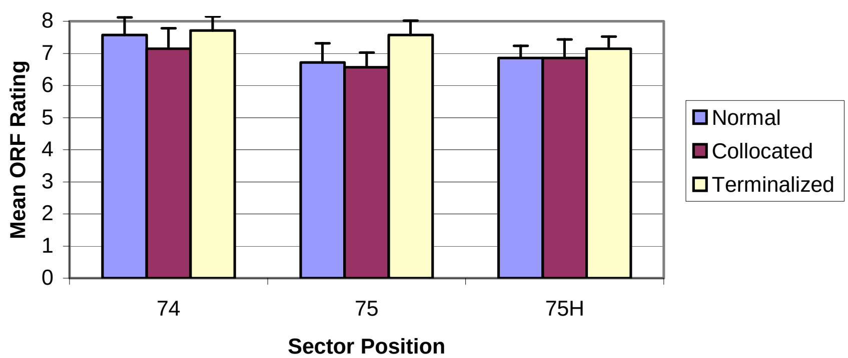


Figure 31. Mean en route ORF ratings for overall providing control information by Sector Position and Condition.

We found significant main effects of both Condition and Sector Position, $F(2, 12) = 4.50$ and $F(2, 12) = 7.36$, respectively, for the participants' demonstrated knowledge of letters of agreement (LOAs) and SOPs. The post hoc tests were not significant. However, the data trend suggests that mean ratings were higher for the Terminalized condition and when the participants were working at Sector 74 (see Figure 32). The mean ratings were very good to most effective in all conditions and sector positions.

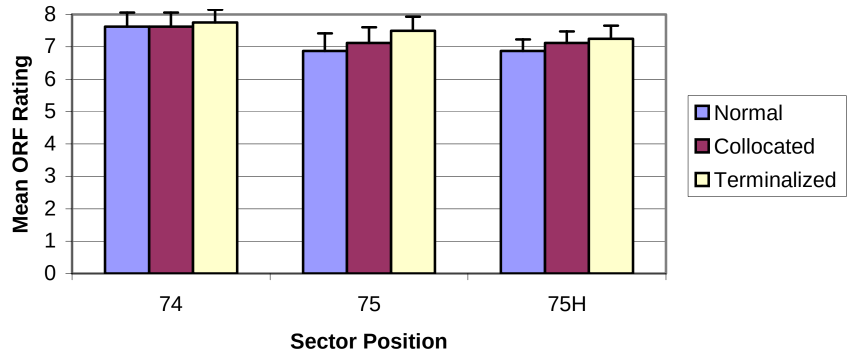


Figure 32. Mean en route ORF ratings for showing knowledge of LOAs and SOPs by Sector Position and Condition.

There was a significant main effect of Sector Position for the participants' ability to share knowledge of aircraft capabilities and limitations. The post hoc test was not significant. However, the participants received the highest ratings when they worked Sector Position 74 (see Figure 33).

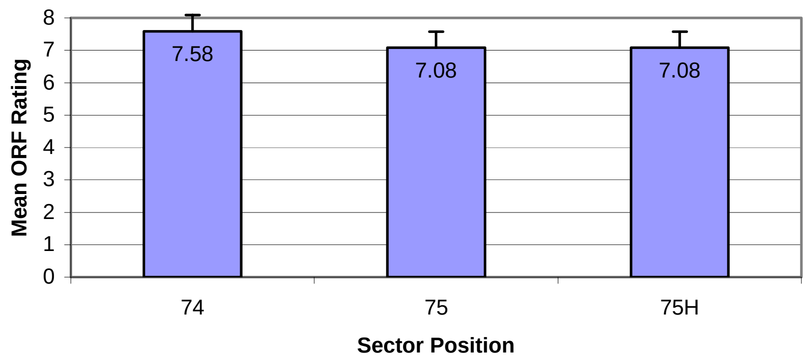


Figure 33. Mean en route ORF ratings for showing knowledge of aircraft capabilities and limitations by Sector Position.

For overall technical knowledge, there was a significant main effect of both Condition and Sector Position, $F(2, 14) = 7.24$ and $F(2, 14) = 13.09$, respectively. The post hoc test for Condition was not significant. However, the participants were rated as displaying more overall technical knowledge in the Terminalized condition. The post hoc test for Sector Position was significant, $HSD(14) = 0.56$, indicating that ratings for Sector Position 74 were higher than ratings for Sector Position 75H (see Figure 34).

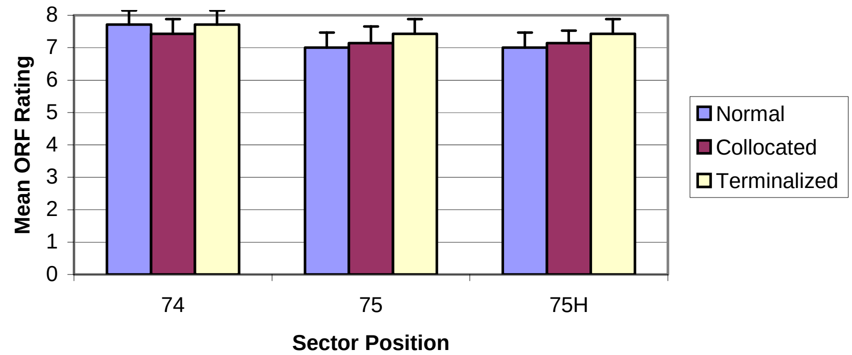


Figure 34. Mean en route ORF ratings overall technical knowledge by Sector Position and Condition.

There was a significant main effect of Condition, $F(2, 16) = 11.58$, for the SME ratings of participants' use of proper phraseology. The participants were more likely to use proper phraseology in the Terminalized condition compared to the Normal condition, HSD(16) = 0.53 (see Figure 35).

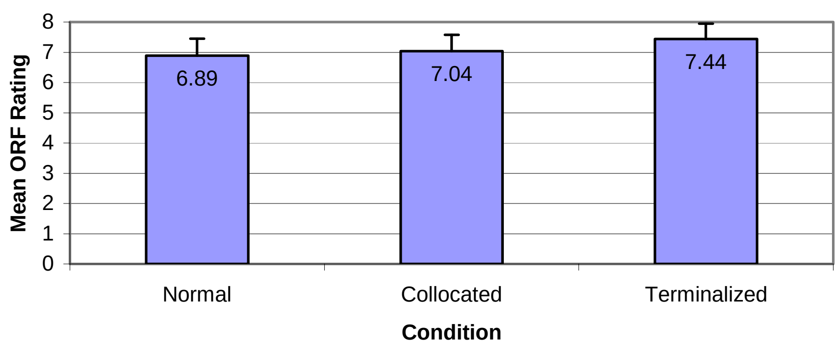


Figure 35. Mean en route ORF ratings for using proper phraseology by Condition.

There was a significant main effect of Condition for the SME ratings of the participants' ability to communicate clearly and efficiently, $F(2, 16) = 10.39$. The post hoc test only revealed a marginally significant difference between the Terminalized and Normal conditions. The general trend for these data suggest that communication was clearest and most efficient during the Terminalized condition (see Figure 36).

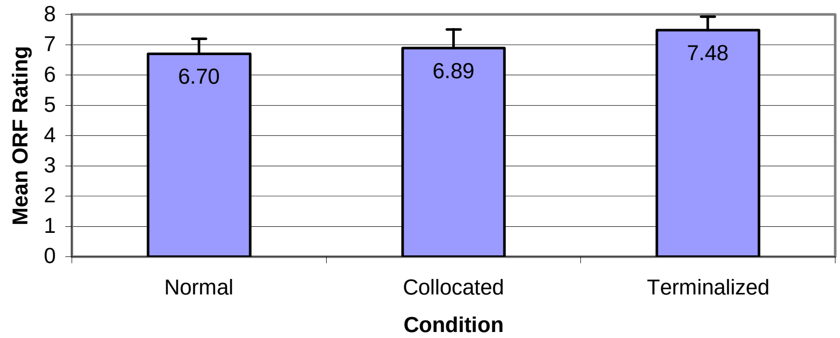


Figure 36. Mean en route ORF ratings for communicating clearly and efficiently by Condition.

There was a significant main effect of Condition for the SME ratings of participants' performance in terms of listening to pilot read backs and requests, $F(2, 16) = 8.08$. The post hoc test only identified a marginal effect between conditions suggesting that performance was higher in the Terminalized condition than in the Normal condition (see Figure 37).

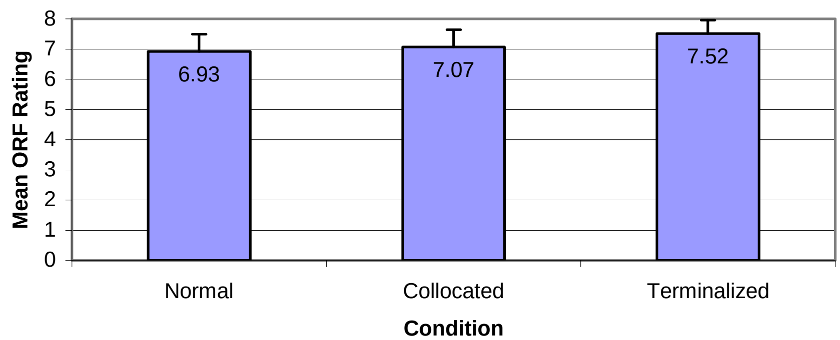


Figure 37. Mean en route ORF ratings listening to pilot read backs and requests by Condition.

For overall communicating, there was a significant main effect of Condition, $F(2, 16) = 8.08$. Again, the post hoc test only revealed a marginal effect pointing towards better overall communication performance in the Terminalized condition compared to the Normal condition (see Figure 38).

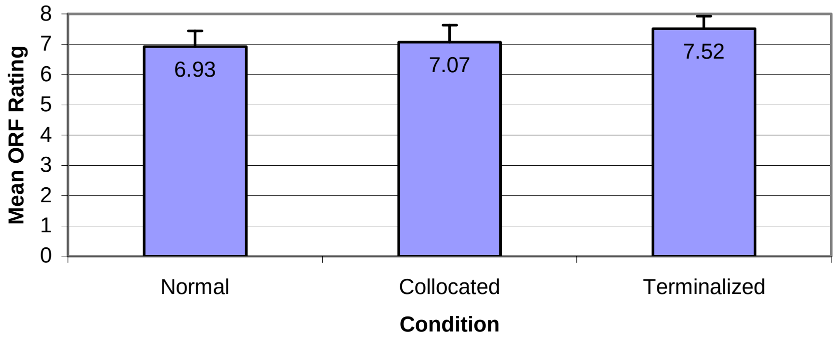


Figure 38. Mean en route ORF ratings overall communicating by Condition.

Overall, from the viewpoint of the SMEs, performance was high at all sector positions and conditions. There were either no differences in participants' performance between conditions or performance was best in the Terminalized condition. For statistically significant effects, the SMEs' subjective ratings of the participants' performance always favored the Terminalized condition. Although the SMEs were aware of which condition we were testing for any given simulation run, they had no reason to be biased in favor of one condition over any other condition. Therefore, for en route participants, we conclude that the Terminalized condition provided benefits in terms of the participants' ability to plan and take control actions, marking FPSs, prioritizing, providing and sharing information, and communication.

3.5.2 Terminal Observer Rating Form

We initially analyzed each item of the terminal ORF using a 3 (Sector Position) X 3 (Condition) repeated measures ANOVA. However, there was a lot of missing data for the ARD H sector position. Because of the quantity of missing data, it was not possible to employ data replacement techniques. Therefore, we dropped the ARD H position from the analyses and performed a 2 (Sector Position; ARD and EWR only) X 3 (Condition) repeated measures ANOVA for each ORF item. None of the analyses detected any significant differences between either Sector Position or Condition. Unlike the en route ORF ratings where even small differences were statistically significant, the terminal ORF ratings contained little or no variability and hence, even small differences between means were either not present or undetectable. Overall, the SMEs rated the participants' performance as high across all sector positions and conditions.

3.6 Workload Measures

The participants provided subjective measures of workload by responding to the WAK and NASA-TLX.

3.6.1 Workload Assessment Keypad Ratings

We analyzed the mean WAK ratings separately for en route and terminal participants. For the analysis of each data set, we used a 3 (Experimental Condition) X 3 (Sector Position) repeated measures ANOVA. If a participant did not respond to a WAK prompt, we treated those data as missing. We chose to treat the WAK data in this way for two reasons. First, we took multiple

WAK ratings (once every 5 min) during a scenario and then averaged these ratings to obtain the unit of analysis. Therefore, we still had a set of observations for the analysis without replacing any data. Second, it is often times not clear why a participant did not respond to a WAK prompt. The participant may have been too busy to respond (suggesting a default rating of 10 would be appropriate) or he may have been using the landline, talking to a pilot, or his attention otherwise diverted such that he never noticed the prompt. Whatever the reason for not responding, it was clear that a participant's failure to respond to the WAK was not always due to being busy. Given the nature of the experiment, it was also likely that a participant was not even near his WAK when the prompt occurred. It is for these reasons that we decided not to assign the maximum rating of 10 to replace missing data.

3.6.1.1 En Route Workload Assessment Keypad Ratings

For the en route WAK ratings, there was a significant main effect of Sector Position, $F(2, 16) = 11.97$. The post hoc test identified a significant difference between Sector Positions 75 and 75H, HSD(16) = 1.56. When the participants were working at Sector Position 75, they rated workload as being significantly higher than when they worked at Sector Position 75H. The participants did not report any changes in subjective workload as assessed by the WAK across conditions. Overall, the en route participants' subjective ratings of workload were at moderate levels despite the relatively high traffic load (130% of normal operations). Figure 39 presents the en route WAK data.

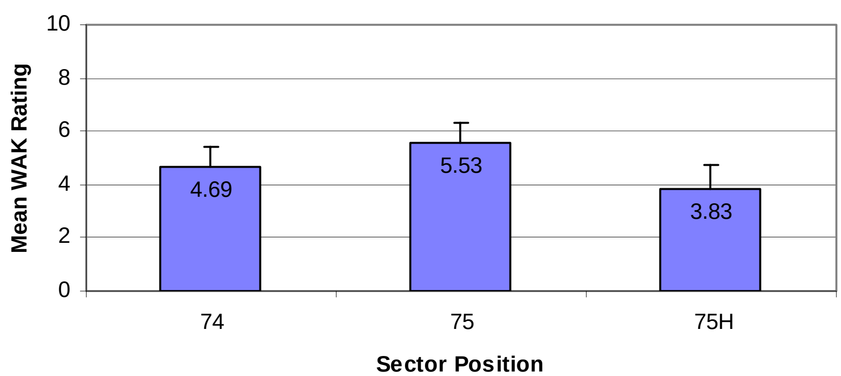


Figure 39. Mean en route WAK ratings by Sector Position.

3.6.1.2 Terminal Workload Assessment Keypad Ratings

For the terminal WAK ratings, there were significant main effects of Sector Position and Condition, $F(2, 16) = 8.93$ and $F(2, 16) = 7.90$, respectively. The post hoc tests were not significant. However, there was a marginal effect ($p < .10$) for sector position, indicating that the mean WAK ratings were higher at EWR than at ARD H. The WAK ratings were lowest in the Normal condition (see Figure 40). In all sector positions and conditions, the WAK ratings indicated low to moderate levels of workload.

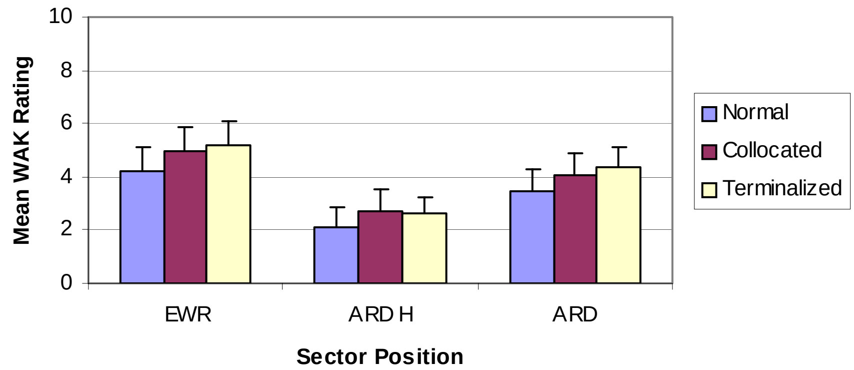


Figure 40. Mean terminal WAK ratings by Sector Position and Condition.

3.6.2 NASA Task Load Index

We analyzed each item of the NASA-TLX separately using a 3 (Experimental Condition: Normal, Collocated, Terminalized) X 3 (Sector Position: EWR, ARD, ARD H for terminal; 74, 75, 75H for en route) repeated measures ANOVA. Items on the NASA-TLX asked participants about their mental demand, physical demand, temporal demand, effort, frustration, and performance. We present the results for the en route and terminal sector positions in turn.

3.6.2.1 En Route NASA-Task Load Index

For the en route participants, two items of the NASA-TLX, mental demand and physical demand, showed significant differences. For the mental demand, there was a significant main effect of Sector Position, $F(2, 16) = 7.01$. The post hoc test did not identify any significant differences. However, the data trend suggests that the participants perceived mental demand as being lowest at Sector Position 75H and somewhat higher at Sector Position 75 (see Figure 41). In all positions, the perceived mental demand was in the moderate to slightly high range.

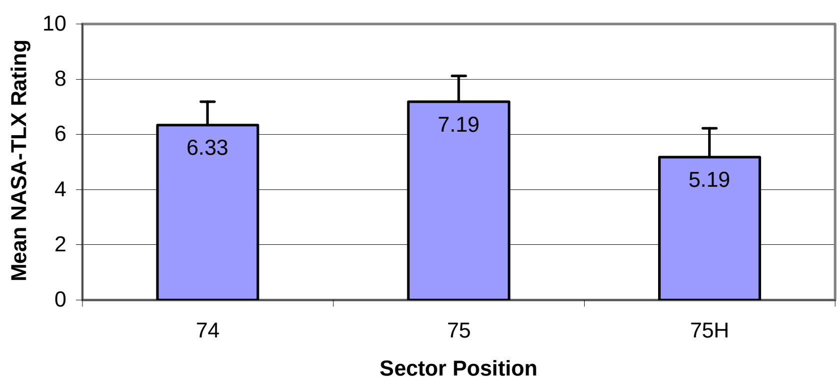


Figure 41. Mean en route NASA-TLX ratings for mental demand by Sector Position and Condition.

For physical demand, there was also a significant main effect of Sector Position, $F(2, 16) = 5.29$. Again, the post hoc test did not identify any significant differences between the sector positions. The data trend indicates that the participants rated Sector Position 75H as being relatively lower in terms of physical demand than either Sector Position 74 or 75 (see Figure 42). The average ratings for physical demand were in the moderate range to slightly high range for all positions.

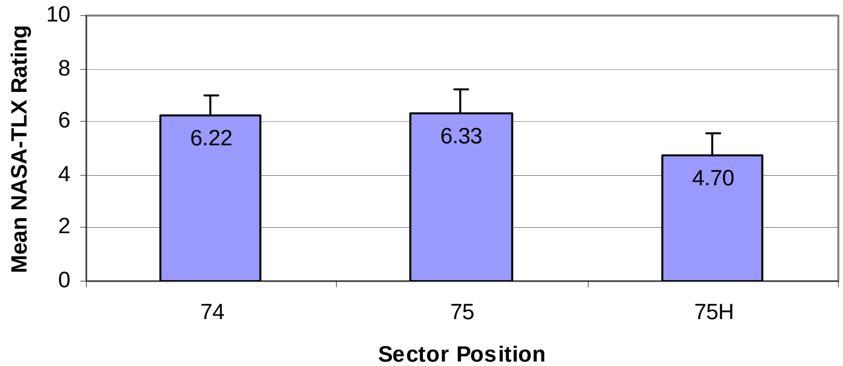


Figure 42. Mean en route NASA-TLX ratings for physical demand by Sector Position and Condition.

We did not find any significant differences for the NASA-TLX items that the en route participants used to rate their level of temporal demand, effort, frustration, or performance. The ratings for temporal demand ranged from 5.7 to 6.4. The participants' ratings of their performance ranged from 7.9 to 8.5. Their ratings of effort ranged from 6.6 to 6.9 and their ratings of frustration ranged from 2.9 to 3.9.

3.6.2.2 Terminal NASA- Task Load Index

For the terminal participants, the analyses identified significant differences in all of the NASA-TLX items except for performance. For mental demand, there was a significant main effect of Sector Position, $F(2, 16) = 10.44$. Mental demand was significantly lower for ARD H than both ARD and EWR, $HSD(16) = 2.54$. For all positions, the rated mental demand ranged from slightly low to slightly high (see Figure 43).

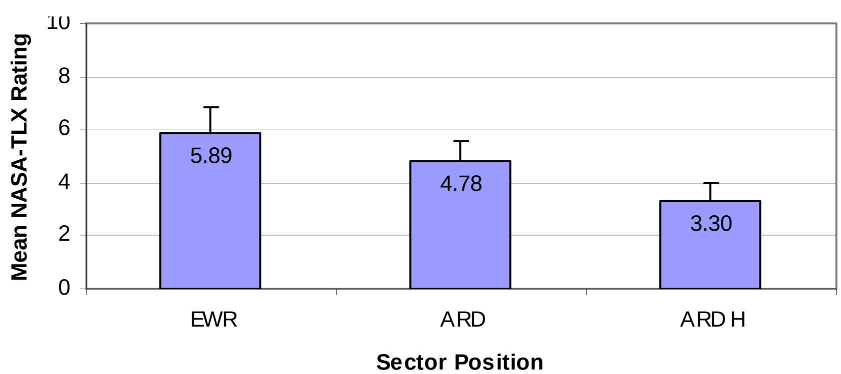


Figure 43. Mean terminal NASA-TLX ratings for mental demand by Sector Position.

For physical demand, there was a significant main effect of Sector Position, *F*(2, 16) = 10.02, and Condition, *F*(2, 16) = 5.52. The post hoc test for sector position was marginal, suggesting that physical demand was lower at the ARD H sector position than both the ARD and the EWR sector positions. The post hoc test for Condition was not significant. However, the data trend suggests that physical demand may have increased slightly for the Collocated and Terminalized conditions compared to the Normal condition, but the average ratings were low to moderate in all conditions (see Figure 44).

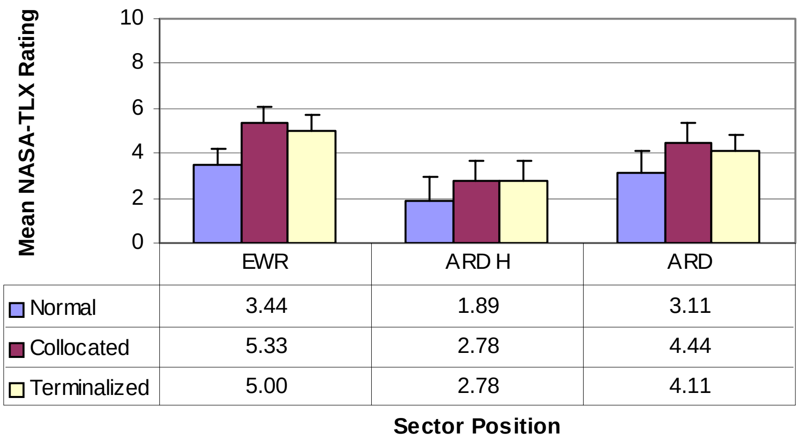
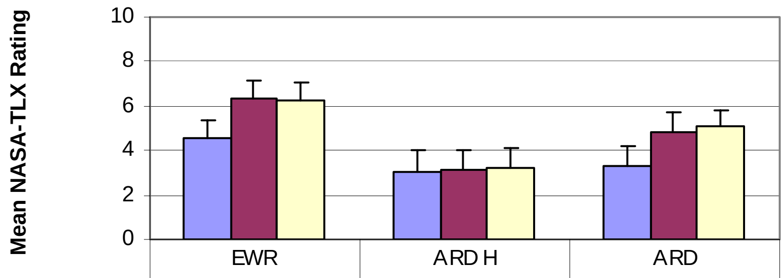


Figure 44. Mean terminal NASA-TLX ratings for physical demand by Sector Position and Condition.

The next item asked participants about their temporal demand. There was a significant main effect of both Sector Position, *F*(2, 16) = 13.03, and Condition, *F*(2, 16) = 6.48, for this item. The post hoc test for sector position was significant, HSD(16) = 2.27, and indicated that temporal demand was significantly lower at the ARD H sector position than either the ARD or EWR sector positions. The post hoc test for Condition was not significant. However, the data trend suggests that, at least for the Sector Positions EWR and ARD, temporal demand was higher during the Collocated and Terminalized conditions compared to the Normal condition (see Figure 45). In those two conditions, the average workload was still low to moderate for ARD and moderate to slightly high for EWR.

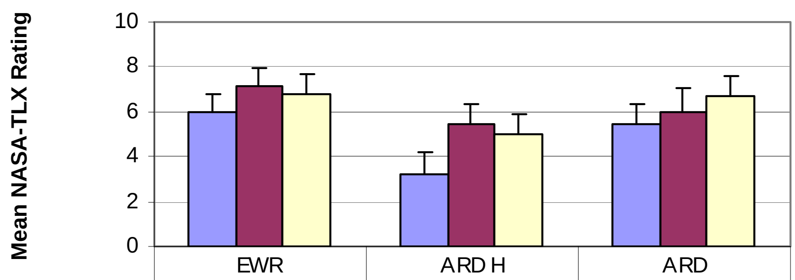
We also found significant main effects of Sector Position, *F*(2, 16) = 5.34, and Condition, *F*(2, 16) = 5.70, for the participants' mean ratings of effort. Neither of the post hoc tests was significant. Although not statistically significant, the data trend suggests that participants may have perceived the Collocated and Terminalized conditions as requiring more effort than the Normal condition and more effort was required at the EWR sector position than at the ARD H sector position (see Figure 46). The mean ratings of effort ranged from low to slightly high.



	EWR	ARD H	ARD
Normal	4.56	3.00	3.33
Collocated	6.33	3.11	4.78
Terminalized	6.22	3.22	5.11

Sector Position

Figure 45. Mean terminal NASA-TLX ratings for temporal demand by Sector Position and Condition.



	EWR	ARD H	ARD
Normal	6.00	3.22	5.44
Collocated	7.11	5.44	6.00
Terminalized	6.78	5.00	6.67

Sector Position

Figure 46. Mean terminal NASA-TLX ratings for effort by Sector Position and Condition.

For the NASA-TLX item that asked participants about their frustration, there was a significant main effect of Sector Position, $F(2, 16) = 7.23$. The post hoc test was not significant. The data trend suggests that the participants perceived frustration to be lowest at the ARD H sector position (see Figure 47).

The participants rated their performance as moderately high throughout the experiment. The mean ratings of performance ranged from 6.33 to 8.33 on the 10-point scale. There were no significant differences in the participants' perception of their performance across either Sector Position or Condition.

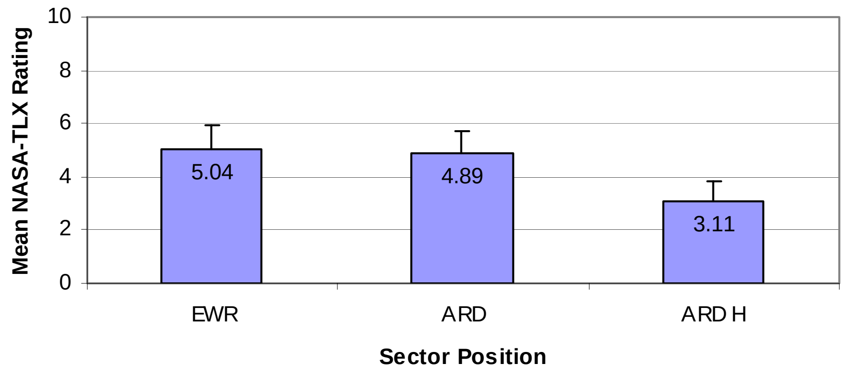


Figure 47. Mean terminal NASA-TLX ratings for frustration by Sector Position.

3.7 Questionnaires

3.7.1 Post-Scenario Questionnaire

We analyzed each item of the PSQ separately using a 3 (Experimental Condition: Normal, Collocated, Terminalized) X 3 (Sector Position: EWR, ARD, ARD H for terminal; 74, 75, 75H for en route) repeated measures ANOVA.

3.7.1.1 En Route Post-Scenario Questionnaire

Table 8 shows all of the means and standard deviations for the PSQ items by Sector Position and Condition. The empty cells of the table indicate that the item was not relevant for that condition. We analyzed each item of the en route PSQ separately using a 3 (Experimental Condition: Normal, Collocated, Terminalized) X 3 (Sector Position: 74, 75, 75H) repeated measures ANOVA.

Table 8. Means and Standard Deviations for the En Route PSQ Items
by Sector Position and Condition

PSQ Item	Sector Position	Normal	Collocated	Terminalized
1. Rate your overall level of ATC performance during this scenario.	74	8.44 (0.73)	8.33 (1.41)	9.11 (0.60)
	75	7.89 (1.17)	8.67 (0.71)	7.67 (1.80)
	75H	8.00 (1.73)	8.56 (1.51)	8.56 (1.81)
2. Rate your ability to move aircraft through the sector during this scenario.	74	8.78 (0.67)	8.33 (1.12)	9.00 (1.22)
	75	7.89 (1.36)	8.56 (0.73)	7.89 (1.83)
	75H	7.78 (1.64)	8.11 (1.90)	8.33 (2.12)
3. Rate your overall level of SA during this scenario.	74	8.44 (1.24)	8.56 (1.42)	8.89 (0.78)
	75	8.11 (1.17)	8.44 (1.01)	7.11 (1.76)
	75H	8.11 (1.45)	8.44 (1.67)	9.11 (0.78)
4. Rate your SA for current aircraft locations during this scenario.	74	8.56 (0.88)	8.44 (1.67)	9.11 (0.78)
	75	7.78 (1.30)	8.67 (0.71)	7.56 (1.88)
	75H	8.22 (1.09)	8.33 (1.87)	9.00 (0.71)
5. Rate your SA for projected aircraft locations during this scenario.	74	8.78 (0.83)	8.33 (1.32)	9.11 (0.78)
	75	8.00 (1.00)	8.44 (0.73)	7.89 (2.03)
	75H	7.78 (1.48)	8.33 (1.50)	8.78 (1.20)
6. Rate your SA for potential aircraft loss-of-separation during this scenario.	74	8.78 (0.97)	8.11 (1.69)	9.22 (0.83)
	75	8.22 (1.64)	9.00 (0.50)	8.00 (1.87)
	75H	8.44 (1.13)	8.89 (1.27)	8.78 (1.09)
7. Rate your workload due to air-to-ground communications during this scenario.	74	6.89 (2.80)	7.44 (2.19)	6.89 (1.76)
	75	6.56 (2.74)	7.11 (1.83)	6.78 (2.05)
	75H	4.78 (3.15)	5.22 (3.38)	4.78 (2.91)
8. Rate your workload due to ground-to-ground communications during this scenario.	74	4.78 (2.86)	2.67 (1.12)	2.67 (1.50)
	75	3.00 (2.55)	3.11 (2.62)	2.67 (1.66)
	75H	4.67 (3.00)	4.11 (2.67)	4.56 (2.83)
9. Rate the performance of the simulation pilots in terms of their responding to your control instructions and providing readbacks.	74	6.44 (1.81)	6.56 (2.01)	7.56 (1.33)
	75	6.78 (2.11)	5.78 (2.86)	7.56 (1.42)
	75H	6.56 (1.67)	6.22 (2.39)	7.33 (1.41)
10. Rate the difficulty of this scenario.	74	6.11 (1.69)	7.00 (1.00)	6.56 (1.59)
	75	6.67 (2.35)	6.89 (1.27)	7.00 (2.35)
	75H	6.78 (2.22)	6.78 (2.17)	7.11 (2.47)
11. What effect, if any, did the reduced lateral separation standards have on your ability to control traffic.	74	…	…	7.44 (1.33)
	75	…	…	7.50 (0.71)
	75H	…	…	6.83 (1.60)
12. Were you able to adapt to the reduced lateral separation standards.	74	…	…	6.78 (2.33)
	75	…	…	7.50 (0.71)
	75H	…	…	5.00 (4.00)

Overall, the ratings were moderate to high for all conditions and sector positions. The en route participants responded differently to three of the PSQ items depending upon the Sector Position and Condition. There was a significant Sector Position X Condition interaction for the participants' ratings of their overall level of SA, $F(4, 32) = 3.26$. The data are shown in Figure 48. When in the Terminalized condition, the participants working Sector Position 75 rated overall SA as being lower than did participants at Sector Positions 74 and 75H, HSD(32) = 1.59. With the exception of Sector 75 in the Terminalized condition, the participants' self-ratings of their overall SA was high (mean rating greater than 8) for all Sector Positions and Conditions.

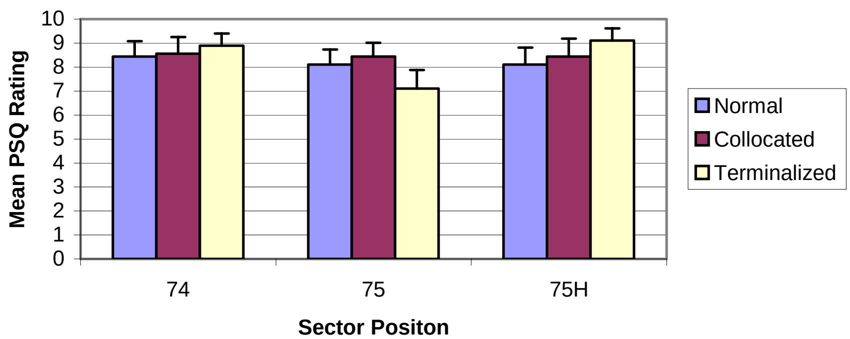


Figure 48. Mean en route PSQ ratings for overall level of SA by Sector Position and Condition.

There was also a significant Sector Position X Condition interaction for participants' ratings of their SA for potential loss of aircraft separation, $F(4, 32) = 2.91$ (see Figure 49). However, the post hoc test was unable to detect any significant differences to explain this interaction. One possible explanation is that in the Terminalized condition, Sector Position 74 rated their SA higher than Sector Position 75. In that condition, Sector Position 74 was using terminal separation standards and was provided a 3-mile J-ring as a tool for maintaining separation.

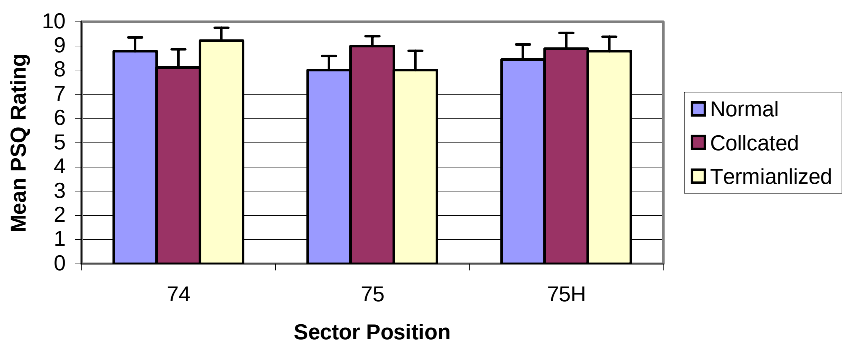


Figure 49. Mean en route PSQ ratings of SA for potential loss of aircraft separation by Sector Position and Condition.

There was a significant main effect of Condition for the participants' ratings of the performance of the simulation pilots, $F(2, 16) = 5.08$, but the post hoc test did not detect any significant differences. The participants perceived that the simulation pilots performed best during the Terminalized condition and lowest during the Collocated condition (see Figure 50).

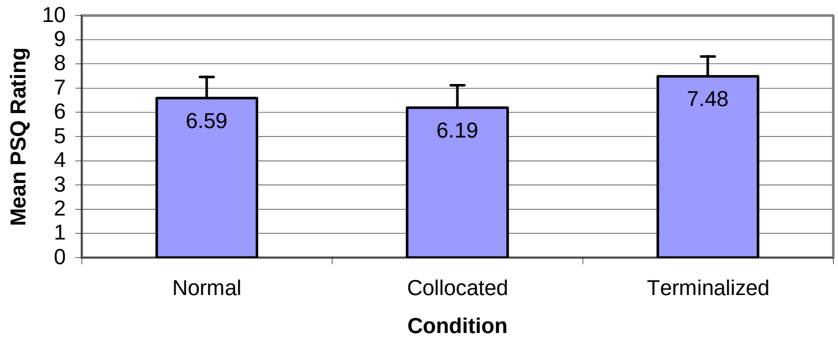


Figure 50. Mean en route PSQ ratings of performance of the simulation pilots by Condition.

After participants completed each scenario under the Terminalized condition, they answered two extra items on the PSQ. Item 11 asked participants, "What effect, if any, did the reduced lateral separations standards have on your ability to control traffic?" Ratings for the PSQ item 11 could range from 1 indicating a *negative effect* to 9 indicating a *positive effect*. A rating of 5 indicated *no effect*. The results for Item 11 appear in Figure 51. We did not perform a statistical test on these data because only two participants at Sector Position 75 responded to the item. Among all en route sector positions, ratings ranged from 5 to 9 with no responses below 5. Therefore, based on the descriptive data alone, we conclude that the majority of participants believed that the reduced lateral separation standards had a positive effect on their ability to control traffic and none believed it had a negative effect.

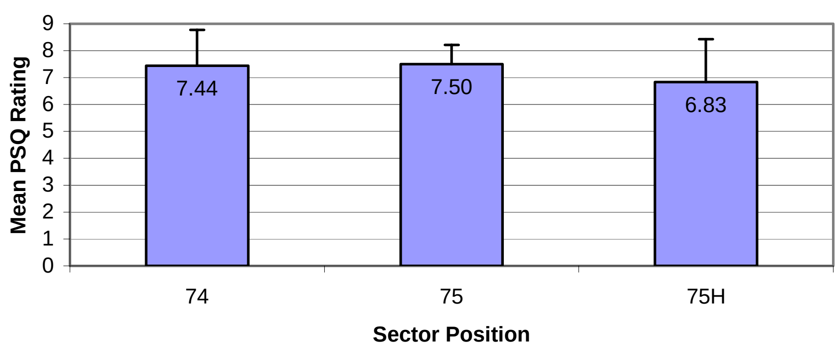


Figure 51. Mean en route PSQ ratings for the effect of reduced lateral separation in the Terminalized condition.

Item 12 asked participants "Were you able to adapt to the reduced lateral separation standards?" The rating scale ranged from 1 indicating participants were *not able to adapt at all* to 10 indicating participants were *able to adapt a great deal*. These data appear in Figure 52. We did not perform a statistical test on these data because of the low number of responses from Sector Positions 75 and 75H. Overall, the participants' mean ratings suggest that they were able to adapt to the reduced lateral separation standards. However, as shown by the error bars, there was a great deal of variability in the ratings for Sector Positions 74 and 75H. For Sector Position 74, participants' ratings ranged from 4 to 10. For Sector Position 75H, participants' ratings ranged from 1 to 9. For Sector Position 75, ratings ranged from 7 to 8. These descriptive statistics suggest that most participants were able to adapt to the new procedure; however, some did have difficulty. The participants at Sector Position 74 were using the reduced separation standards, and the participants at Sector Position 75H were handing off traffic between two sectors using different separation standards. The participants at Sector Position 75 may have been affected somewhat less by the reduced separation standards as they reported that they were the most able to adapt to the Terminalized condition.

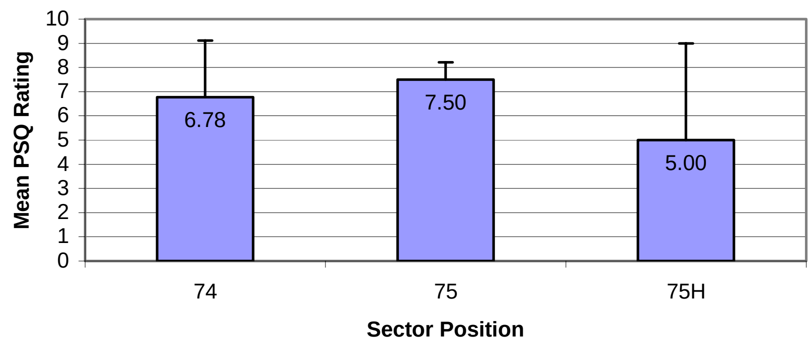


Figure 52. Mean en route ratings for ability to adapt to reduced lateral separation in the Terminalized condition.

3.7.1.2 Terminal Post-Scenario Questionnaire

Table 9 shows all of the means and standard deviations for the PSQ items by Sector Position and Condition. The empty cells of the table indicate that the item was not relevant for those conditions. We analyzed each item of the terminal PSQ separately using a 3 (Experimental Condition: Normal, Collocated, Terminalized) X 3 (Sector Position: EWR, ARD, ARD H) repeated measures ANOVA.

Table 9. Means and Standard Deviations for the Terminal PSQ Items
by Sector Position and Condition

PSQ Item	Sector Position	Normal	Collocated	Terminalized
1. Rate your overall level of ATC performance during this scenario.	EWR	7.22 (1.86)	8.22 (2.33)	7.89 (1.90)
	ARD H	6.67 (2.12)	6.56 (1.88)	6.78 (2.73)
	ARD	7.89 (2.80)	8.22 (1.86)	8.56 (1.59)
2. Rate your ability to move aircraft through the sector during this scenario.	EWR	7.44 (1.74)	8.78 (1.30)	8.44 (1.81)
	ARD H	6.22 (2.49)	7.22 (1.92)	7.22 (3.23)
	ARD	6.78 (2.99)	7.89 (2.20)	9.00 (1.32)
3. Rate your overall level of SA during this scenario.	EWR	7.89 (1.83)	9.00 (1.12)	8.33 (1.58)
	ARD H	7.89 (1.69)	7.11 (1.69)	8.22 (1.86)
	ARD	8.89 (1.36)	8.00 (2.12)	8.67 (1.41)
4. Rate your SA for current aircraft locations during this scenario.	EWR	7.56 (1.59)	8.78 (1.30)	8.56 (1.33)
	ARD H	7.67 (1.58)	7.33 (1.58)	8.11 (1.83)
	ARD	9.11 (1.36)	8.33 (1.87)	8.89 (1.17)
5. Rate your SA for projected aircraft locations during this scenario.	EWR	7.78 (1.39)	8.56 (1.67)	8.11 (1.36)
	ARD H	7.44 (1.51)	7.11 (1.76)	8.00 (2.00)
	ARD	8.89 (1.27)	8.56 (1.94)	8.89 (1.17)
6. Rate your SA for potential aircraft loss-of-separation during this scenario.	EWR	8.11 (1.27)	8.56 (1.88)	8.33 (1.41)
	ARD H	7.89 (1.62)	7.11 (1.90)	7.78 (1.92)
	ARD	9.11 (1.69)	8.11 (2.80)	9.00 (1.12)
7. Rate your workload due to air-to-ground communications during this scenario.	EWR	7.33 (2.55)	6.78 (3.23)	7.33 (2.40)
	ARD H	3.22 (2.44)	5.33 (2.24)	3.78 (2.77)
	ARD	6.22 (2.44)	6.11 (2.52)	6.11 (3.69)
8. Rate your workload due to ground-to-ground communications during this scenario.	EWR	5.67 (3.28)	5.00 (3.74)	5.11 (3.33)
	ARD H	4.11 (1.96)	4.78 (2.05)	4.67 (2.65)
	ARD	4.22 (2.68)	4.00 (2.50)	3.56 (2.65)
9. Rate the performance of the simulation pilots in terms of their responding to your control instructions and providing readbacks.	EWR	7.11 (2.67)	7.89 (2.47)	7.78 (2.68)
	ARD H	6.88 (2.47)	6.33 (2.18)	7.56 (1.81)
	ARD	8.22 (2.33)	8.22 (1.86)	8.22 (2.28)
10. Rate the difficulty of this scenario.	EWR	5.56 (1.67)	6.56 (3.13)	6.89 (1.83)
	ARD H	3.67 (2.29)	5.89 (1.96)	6.89 (2.62)
	ARD	4.11 (2.15)	6.11 (2.03)	6.11 (2.20)
11. What effect, if any, did the reduced lateral separation standards have on your ability to control traffic.	EWR	…	…	6.29 (2.36)
	ARD H	…	…	5.78 (1.56)
	ARD	…	…	7.22 (2.33)
12. Were you able to adapt to the reduced lateral separation standards.	EWR	…	…	8.00 (2.45)
	ARD H	…	…	7.33 (1.97)
	ARD	…	…	8.44 (1.88)

Overall, the PSQ ratings were moderate to high for all conditions and sector positions. The PSQ Items 2, 3, 7, and 10 showed statistically significant effects for the terminal participants. There was a significant main effect of Condition for participants' ratings about their ability to move aircraft through their Sector, $F(2, 16) = 4.30$. The post hoc test did not detect any significant differences. However, the data trend suggests that participants were least able to move aircraft through their sector during the Normal condition and best able to move aircraft during the Terminalized condition (see Figure 53).

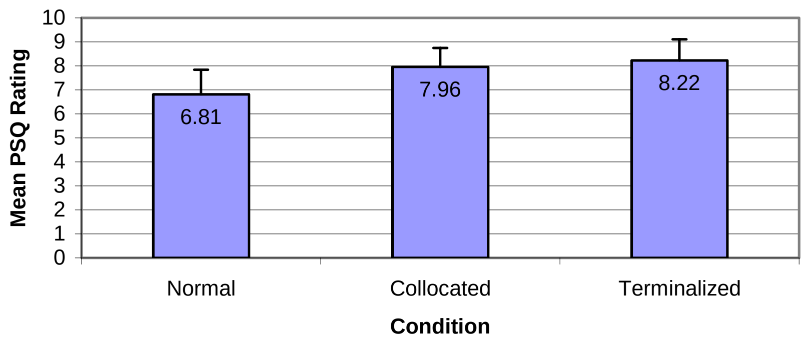


Figure 53. Mean terminal PSQ ratings for ability to move aircraft through the Sector by Condition.

There was a significant Sector Position X Condition interaction for the participants' ratings of their overall level of SA, $F(4, 32) = 3.51$ (see Figure 54). The post hoc test revealed that during the Collocated condition, participants at ARD H rated their overall SA as lower than when they worked the EWR sector position, HSD(32) = 1.54. The post hoc test further indicated a lower participant rating for the ARD H sector during the Collocated condition compared to the ARD sector during both the Normal and Terminalized conditions. The participants' subjective ratings were moderately high to high in all positions and conditions, suggesting that they believed they had adequate to near optimal levels of SA throughout the experiment.

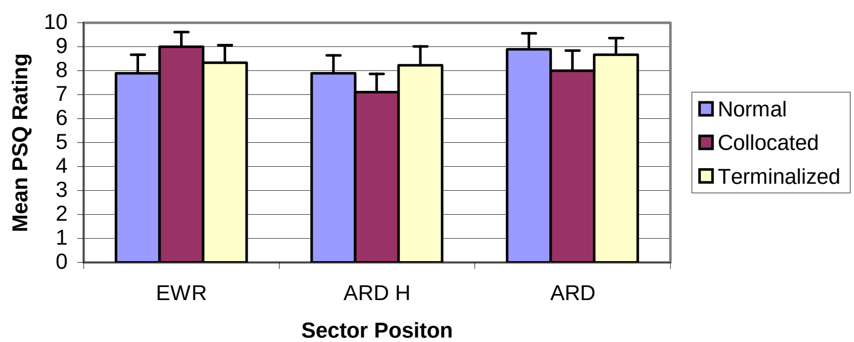


Figure 54. Mean terminal ratings for overall level of SA by Sector Position and Condition.

There was a significant main effect of sector position for participants' ratings of their workload resulting from air-ground communications, $F(2, 16) = 6.64$. The post hoc test was not significant. However, as one would expect, the data trend suggests that workload due to air-ground communications was lowest at the handoff position and somewhat higher at ARD and EWR where the participants were talking to aircraft almost exclusively (see Figure 55).

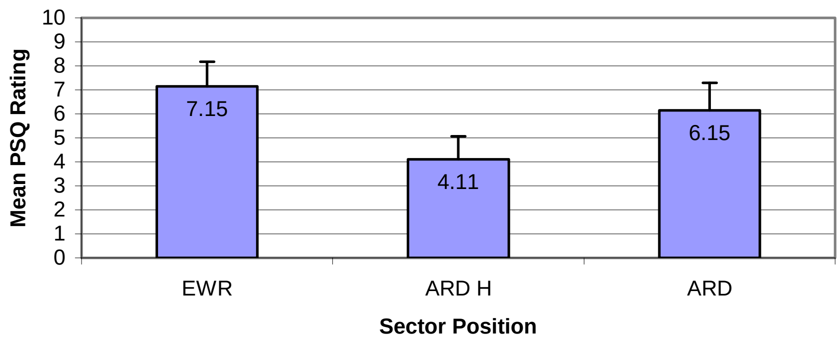


Figure 55. Mean terminal PSQ ratings for workload due to air-ground communications by Sector Position.

Item 10 of the PSQ asked participants to rate the scenario difficulty. There was a significant main effect of both Condition and Sector Position, $F(2, 16) = 15.26$ and $F(2, 16) = 5.25$, respectively. Figures 56 and 57 display the data. The post hoc test indicated that the participants perceived the Terminalized condition as significantly more difficult than the Normal condition, HSD(16) = 1.87. Although the participants rated the scenarios as being more difficult during the Terminalized condition, this perceived difference was not likely due to inherent differences between the scenarios themselves. The SMEs created the scenarios to be as similar as possible. Furthermore, we controlled for any unforeseen differences in scenarios by counterbalancing the order and condition in which participants experienced the scenarios. The participants may have rated the scenarios in the Collocated and Terminalized conditions as being slightly more difficult because they were moving more aircraft and were experiencing higher levels of taskload and workload. The post hoc test for sector position was not significant. The participants may have been exposed to relatively more pressure when they worked the EWR sector because they tended to rate the scenarios as being more difficult than either ARD or ARD H.

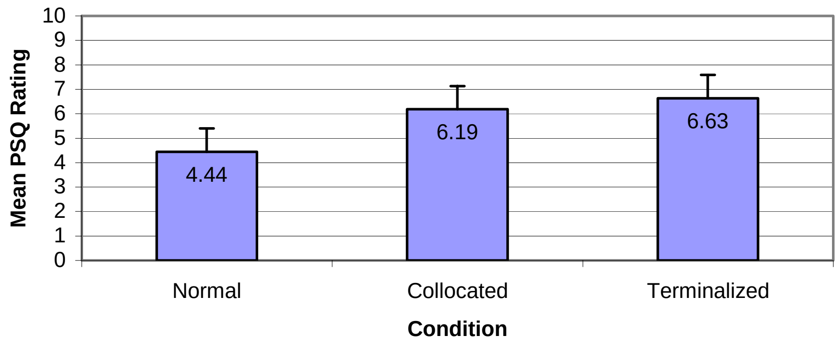


Figure 56. Mean terminal PSQ ratings for scenario difficulty by Condition.

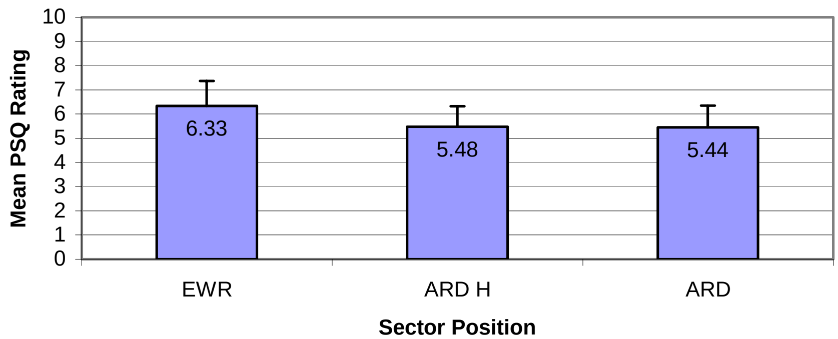


Figure 57. Mean terminal PSQ ratings for scenario difficulty by Sector Position.

The terminal participants answered two extra items on the PSQ after each scenario was completed under the Terminalized condition. Item 11 asked participants, "What effect, if any, did the reduced lateral separation standards have on your ability to control traffic?" Ratings for this item could range from 1 indicating a *negative effect* to 9 indicating a *positive effect*. A rating of 5 would indicate *no effect*. Based on the descriptive data, we can infer that the terminal participants rated the reduced lateral separation standards as having a somewhat positive effect on their ability to control traffic (see Figure 58). The participants tended to perceive a greater positive benefit of the Terminalized condition when they worked at the ARD sector position, but there were no statistically significant differences between positions.

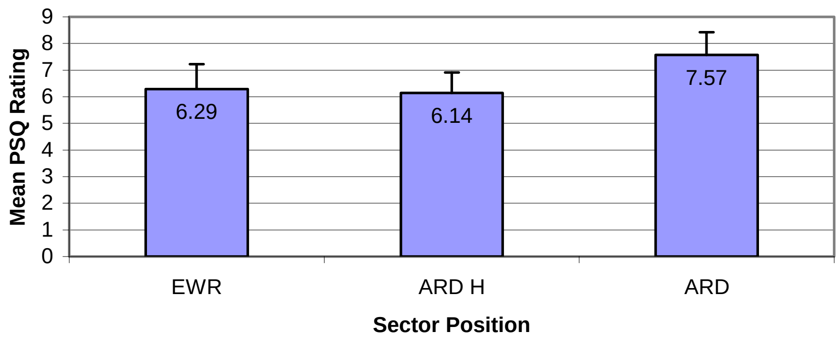


Figure 58. Mean terminal PSQ ratings for the effect of reduced lateral separation in the Terminalized condition by Sector Position.

Item 12 asked, "Were you able to adapt to the reduced lateral separation standards?" The rating scale ranged from 1 indicating participants were *not able to adapt at all* to 10 indicating participants were *able to adapt a great deal*. Overall, participants' mean ratings suggest that they were able to adapt and use the new procedure (see Figure 59). We did not find any significant differences between sector positions and their ability to adapt to the new procedure, but the ratings were highest in ARD, which was interacting with Sector 74, which was using the terminal separation standards.

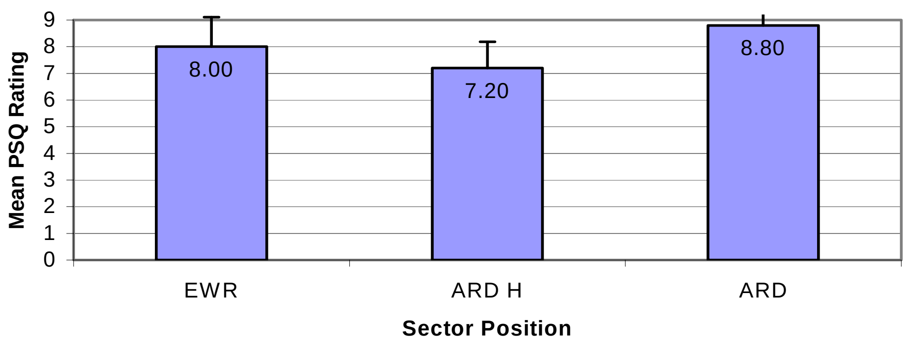


Figure 59. Mean terminal PSQ ratings for ability to adapt to reduced lateral separation in the Terminalized condition.

3.7.2 Post-Experiment Questionnaire

All participants completed the PEQ at the end of the experiment. The participants rated Items 1 and 3 of the PEQ using a 10-point scale where a rating of 1 indicated *no change at all* and a 10 indicated *a great deal of change*. The participants rated Items 2 and 4 using a 9-point scale where a rating of 1 indicated a *negative effect*, a rating of 9 indicated a *positive effect*, and a rating of 5 indicated *no effect*. Table 10 shows the ratings from the en route and terminal participants, respectively.

Table 10. Means and Standard Deviations of En Route and Terminal Participants' PEQ Responses

	En Route – ZNY	Terminal – N90
PEQ Item	**Mean (*SD*)**	**Mean (*SD*)**
1. Did your communication strategies change during the Collocated Condition?	9.50 (1.41)	8.22 (2.91)
2. What effect, if any, did collocation alone have on your control strategies?	7.50 (1.77)	8.22 (0.97)
3. Did your communication strategies change during the Terminalized Condition?	8.00 (2.83)	7.56 (3.36)
4. What effect, if any, did the Terminalized Condition have on your control strategies?	8.00 (1.26)	7.67 (1.73)

Both the en route and terminal participants' PEQ ratings indicate that both the Collocated and Terminalized conditions had a high positive effect on their communication and control strategies.

4. CONCLUSION – ARRIVAL EXPERIMENT

In comparison to the Normal condition, the system performance measures for the Terminalized condition showed an increase in the number of aircraft handled, an increase in the number of completed flights (aircraft handed off to the tower), a decrease in the number of holds, a decrease in the duration of holds, and a stable number of control commands. The Collocated condition tended to show some benefits as well, but not to the same degree as the Terminalized condition. The Collocated condition had a greater number of holds than the Normal condition, but they were of shorter duration. For these measures, the Collocated condition may have increased taskload somewhat.

The Terminalized condition significantly reduced the number of losses of separation in Sector 74 compared to the Normal condition. The EWR sector had significantly more losses of separation and wake vortex violations than the ARD sector, but there were no differences across conditions. All of the safety measures probably overestimated the rate of occurrence because we could not evaluate every possible instance to make sure it was an actual violation.

For the communication measures, the Collocated and Terminalized conditions decreased the number and duration of landline transmissions between ZNY and N90 compared to the Normal condition, whereas interfacility landline transmissions were more frequent and of a longer duration. That is, when the participants were able to use FTF communication between ZNY and N90, there were more landline transmissions with ZDC, but not with ZOB. The number and duration of ground-air transmission was relatively stable across the sector positions and conditions. The participants compensated for the reduced landline activity by engaging in more FTF communication during the Collocated and Terminalized conditions. The ARD participants in particular took advantage of the collocated situations by obtaining information from the adjacent en route display of Sector 74.

The data from the ORF indicated that all the participants performed well throughout the experiment. The en route participants received the highest mean ratings in the Terminalized condition for 12 different items of the ORF. The SMEs rated the en route participants as performing best in the Terminalized condition in terms of (a) taking actions in appropriate order of importance, (b) preplanning for control actions, (c) marking FPSs while performing other actions, (d) prioritizing overall, (e) providing control information overall, (f) showing knowledge of LOAs and SOPs, (g) showing knowledge of aircraft capabilities and limitations, (h) overall technical knowledge, (i) using proper phraseology, (j) communicating clearly and efficiently, (k) listening to pilot read backs and requests, and (l) communicating overall. The SMEs' subjective ratings of the terminal participants' performance did not indicate any significant differences between conditions, but they did not identify any areas of concern either.

The participants indicated on the PSQ that the reduced lateral separation standards used in the Terminalized condition had a general positive effect on their ability to control air traffic. The en route participants were able to adapt to the new procedures and use them effectively despite never having used the reduced separation procedures prior to this experiment. The subjective rating data from the PSQ showed that the en route participants did not believe that the Collocated or Terminalized conditions affected them negatively. Their overall subjective ratings of their SA were high. The data from the PSQ also showed that terminal participants thought they were best able to move aircraft through the airspace during the Terminalized condition. Both terminal and en route participants reported that they were able to adapt to the new procedures and that the reduced lateral separation standard had a positive effect on their ability to control traffic. The participants reported on the PEQ that both the Collocated and Terminalized conditions positively affected their communication and control strategies. They rated the realism of the simulation as moderate to high and indicated that interference from the WAK device was negligible.

For en route participants, the subjective ratings of workload were moderate and did not change across conditions. For the terminal participants, WAK ratings were low to moderate and tended to be lower in the Normal condition and higher in the Terminalized condition. The difference between WAK ratings across conditions was not statistically significant though.

5. METHOD – DEPARTURE EXPERIMENT

5.1 Participants

Each participant read and signed an informed consent statement prior to the experiment. They also completed a biographic questionnaire. The experimenters kept participant names separate from data to ensure confidentiality. The participants did not experience unusual stress during the experiment and were not at risk.

Thirty-two male CPCs participated in the departure simulation. Eight of the CPCs were from ZNY en route Sectors Parke (39) and Lanna (55); 24 were from N90 terminal Sectors Liberty West (LIBW), EWR, and LGA. The en route CPCs participated in two groups of four. The terminal CPCs participated in four groups of six. Twelve of the participants, three from ZNY and nine from N90, wore corrective lenses during the experiment. Table 11 shows the descriptive statistics (means and *SD*s) for the biographic questionnaire data provided by the en route and terminal CPC participants.

Table 11. Biographic Questionnaire Means and Standard Deviations for the En Route CPC Participants

	En Route - ZNY	**Terminal – N90**
Biographic Questionnaire Item	**Mean (*SD*)**	**Mean (*SD*)**
What is your age?	39.63 (4.98)	36.96 (5.95)
How long have you worked as a CPC (include both FAA and military experience)?	15.82 (4.01)	14.01 (5.19)
How long have you worked as a CPC for the FAA?	15.82 (4.01)	11.07 (5.44)
How long have you been a Certified Professional Controller (or Full Performance Level Controller)?	13.10 (3.24)	11.41 (6.07)
How long have you actively controlled traffic in the en route/terminal environment?	14.50 (3.77)	12.91 (5.71)
How many of the past 12 months have you actively controlled traffic?	11.99 (0.35)	12.00 (0.00)
Rate your current skill as a CPC.	9.00 (0.76)	9.21 (0.93)
Rate your current level of stress.	4.75 (1.83)	3.42 (2.10)
Rate your level of motivation to participate in this study.	8.38 (1.60)	8.08 (2.17)

Overall, the participants were highly experienced CPCs. The en route participants were somewhat more experienced than the terminal participants. All of the participants rated their skill level as being high. The participants rated their current stress levels as low to moderate and their motivation to participate in the experiment as moderate to high.

5.2 Research Personnel

Four experimenters, a PI, CPI, and two RAs conducted the experiment. Because there was an additional sector position in this experiment, we added one SME to serve as an over-the-shoulder observer. Approximately 16 simulation pilots assisted during the scenario shakedown and experimental runs. Representatives from the New York Airspace Redesign Team and CPCs from ZNY and N90 participated in the shakedown exercises to ensure the scenarios were realistic. The hardware and software engineers were present as in the arrival experiment.

5.3 Equipment

We conducted this study at the FAA WJHTC RDHFL. We used the same hardware and software for the departure experiment as we used for the arrival experiment with the exception of an additional terminal workstation console (R-side and Handoff positions) and an en route D-side position (as shown in Figure 3). We arranged the terminal sectors with EWR on the left, LGA in the center, and LIBW on the right. The terminal handoff positions were located to the left of the R-side. The en route Sector 39 was to the immediate right of LIBW and Sector 55 was to the right of 39. The en route handoff positions were located on the right of the R-side position. In addition to the simulated sectors, we also simulated departures from JFK, Teterboro (TEB), Westchester County (HPN), and Morristown Municipal (MMU) airports. We also included two Ghost sectors staffed by simulation pilots to simulate adjacent sectors in the ZOB and ZDC ARTCCs.

5.4 Materials

We used the same materials in the departure experiment as we used in the arrival experiment including the Biographic questionnaire, NASA TLX, PSQ, PEQ, CSS, en route ORF, terminal ORF, and SOPs. The SMEs generated the scenarios in the same manner. The SMEs started with three base scenarios and created four versions of each for 12 scenarios in total. The scenarios were each 50 min in length. We also removed all aircraft with a "Heavy" designation from the scenarios so that we could release aircraft from EWR and LGA at the approximate rate of one aircraft per minute from each airport.

5.5 Experimental Design

We essentially conducted two experiments within the same simulation. Because the participants were not qualified to control traffic in both terminal and en route sectors, we treated the data collected within each airspace type separately. For the en route airspace, the experiment used a 3 (Condition: Normal, Collocated, Terminalized) X 2 (Sector: 39, 55) X 2 (Position: Radar, Handoff) repeated measures design such that participants controlled traffic at each en route sector and position. While at each position, the en route participants experienced each of the three conditions (Normal, Collocated, Terminalized) once. The experimental design for the terminal airspace was different because participants were only certified to control traffic at one of the simulated sectors. Therefore, for the terminal airspace, we used a 3 (Condition: Normal, Collocated, Terminalized) X 3 (Sector: WEST, EWR, LGA) X 2 (Position: Radar, Handoff) mixed effects design with sector being a between-subjects factor, and Position and Condition being a within-subjects factor. While at each sector position, the participants experienced each of the three conditions (Normal, Collocated, Terminalized) once in succession. Table 12 displays the counterbalancing scheme that we used for the conditions and scenarios.

5.5.1 Independent Variables

During the Normal condition, the participants controlled traffic as they normally would in the field, and a wall physically separated the terminal and en route sectors. During the Collocated condition, the experimenters removed the wall from between the terminal and en route sectors making FTF communication between the sectors possible. During the Terminalized condition, we removed the wall (i.e., the Collocated condition) and we reduced the lateral separation standard for en route Sectors 39 and 55 (adjacent to terminal Sector LIBW) from 5 nm to 3 nm. We also reduced the radius of the J-ring from 5 nm to 3 nm.

Table 12. Departure Experiment Condition and Scenario Counterbalancing

		En Route Sector				Terminal Sector							
Group	Simulation Run	55	55H	39	39H	LIBW	LIBW H	EWR	EWRH	LGA	LGAH	Condition	Scenario
1-1	1	E1	E2	E3	E4	T1	T2	T3	T4	T5	T6	N	3A
"	2	"	"	"	"	"	"	"	"	"	"	C	1A
"	3	"	"	"	"	"	"	"	"	"	"	T	2A
"	4	E2	E3	E4	E1	T2	T1	T4	T3	T6	T5	C	3B
"	5	"	"	"	"	"	"	"	"	"	"	T	1B
"	6	"	"	"	"	"	"	"	"	"	"	N	2B
1-2	7	E3	E4	E1	E2	T7	T8	T9	T10	T11	T12	T	1C
"	8	"	"	"	"	"	"	"	"	"	"	N	2C
"	9	"	"	"	"	"	"	"	"	"	"	C	3C
"	10	E4	E1	E2	E3	T8	T7	T10	T9	T12	T11	T	1D
"	11	"	"	"	"	"	"	"	"	"	"	C	3D
"	12	"	"	"	"	"	"	"	"	"	"	N	2D
2-3	13	E5	E6	E7	E8	T13	T14	T15	T16	T17	T18	C	2A
"	14	"	"	"	"	"	"	"	"	"	"	N	1A
"	15	"	"	"	"	"	"	"	"	"	"	T	3A
"	16	E8	E5	E6	E7	T14	T13	T16	T15	T18	T17	N	3B
"	17	"	"	"	"	"	"	"	"	"	"	T	2B
"	18	"	"	"	"	"	"	"	"	"	"	C	1B
2-4	19	E7	E8	E5	E6	T19	T20	T21	T22	T23	T24	N	1C
"	20	"	"	"	"	"	"	"	"	"	"	C	2C
"	21	"	"	"	"	"	"	"	"	"	"	T	3C
"	22	E6	E7	E8	E5	T20	T19	T22	T21	T24	T23	T	2D
"	23	"	"	"	"	"	"	"	"	"	"	N	3D
"	24	"	"	"	"	"	"	"	"	"	"	C	1D

(*Note*. E1 = Participant 1 for en route; T1 = Participant 1 for terminal; Condition N = Normal, Condition C = Collocated, Condition T = Terminalized).

5.5.2 Dependent Variables

In general, we used the same dependent variables that we used in the arrival experiment. For each condition, the experimenters and SMEs obtained measures of communication, performance, and efficiency. Because this experiment focused on departures instead of arrivals, we did not measure airborne holding or the number of completed flights. Instead, we examined the number of departures, number and duration of departure stops, duration of departure delays, and departure intervals. The participants provided subjective measures of performance, SA, and workload as in the arrival experiment.

5.6 Procedure

We used the same general procedure for the departure experiment as we used for the arrival experiment. This section details only the differences between the two studies.

5.6.1 General Schedule of Events

The en route participants were involved in the experiment for three days. The terminal participants were involved in the experiment for approximately one and a half days. Table 13 shows the daily schedule of events for each of the two weeks.

Table 13. Daily Schedule of Events

First Day		**Second Day**		**Third Day**	
Time	**Event**	**Time**	**Event**	**Time**	**Event**
8:30	Intro, Informed Consent, Bio	8:30	Equipment Familiarization	8:30	Equipment Familiarization
9:30	Equipment Familiarization	8:50	Break	8:50	Break
10:00	Break	9:00	Exp. Run 5	9:00	Exp. Run 9
10:15	Exp. Run 1	10:00	Break	10:00	Break
11:15	Break	10:15	Exp. Run 6	10:15	Exp. Run 10
11:30	Exp. Run 2	11:15	Out Briefing, Questionnaires for Terminal CPCs	11:15	Break
12:30	Lunch	12:00	Lunch	11:30	Equipment Familiarization
1:30	Exp. Run 3	1:00	Intro, Informed Consent, Bio for Terminal CPCs	11:50	Lunch
2:30	Break	2:00	Equipment Familiarization	1:00	Exp. Run 11
2:45	Equipment Familiarization	2:20	Break	2:00	Break
3:05	Break	2:30	Exp. Run 7	2:15	Exp. Run 12
3:15	Exp. Run 4	3:30	Break	3:15	Break
4:15	Break	3:45	Exp. Run 8	3:30	Exp. Run 13 (if needed)
4:30	Daily Out Briefing	4:45	Daily Out Briefing	4:30	Out Briefing, Questionnaires

5.6.2 Data Collection Procedure

The participants took part in the experiment according to the counterbalancing scheme shown in Table 12. At the beginning of each scenario, the participants saw a countdown screen on their workstation display. The countdown screen was black with white numbers in the center that counted down from 60 to zero. Thus, the countdown screen remained visible for about one minute during which time the air traffic scenario was loading into DESIREE. We implemented the countdown screen primarily to ensure that the numerous computer processors that we were using had time to process the large volume of simulated air traffic. Once the countdown screen disappeared and was replaced by air traffic, the procedure continued as in the arrival experiment.

6. RESULTS – DEPARTURE EXPERIMENT

We analyze and report the data by employing the same approach as in the arrival experiment.

6.1 Simulation Realism

The participants provided their opinion of the simulation realism by responding to six items on the PEQ using a 10-point scale (see Table 14). Overall, the participants rated the realism of the simulation as being moderate to high. The ratings indicated that the simulated airspace was highly realistic, but the hardware and software were only moderately realistic. They also indicated that the WAK device created little interference with their ATC performance.

Table 14. Means and Standard Deviations for the PEQ Items on Simulation Realism

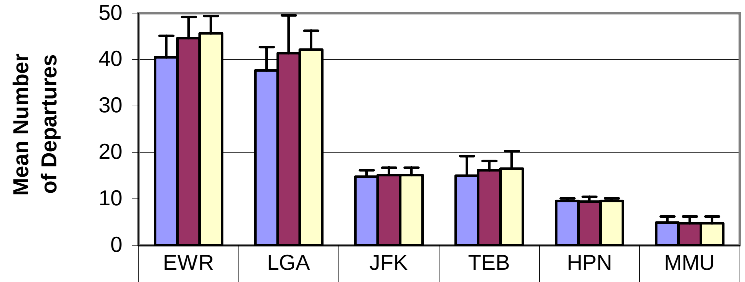
		Terminal – N90	**En Route - ZNY**
Item No.	**PEQ Item**	**Mean (*SD*)**	**Mean (*SD*)**
5	Rate the realism of the overall simulation experience compared to actual ATC operations.	6.33 (1.76)	7.00 (1.41)
6	Rate the realism of the simulation hardware compared to actual equipment.	5.88 (1.78)	7.00 (1.77)
7	Rate the realism of the simulation software compared to actual functionality.	5.96 (1.78)	6.50 (2.14)
8	Rate the realism of the simulation traffic scenarios compared to actual NAS traffic.	6.17 (1.97)	7.25 (2.38)
9	Rate the realism of the simulation airspace compared to actual NAS airspace.	8.96 (1.08)	9.50 (0.76)
10	To what extent did the WAK online workload rating technique interfere with your ATC performance?	1.88 (1.08)	1.50 (1.07)

6.2 System Performance Measures

We collected all of the system data during twenty-four 50 min simulation scenarios. We used the same methodology as the arrival experiment to collect the system performance measures. We describe the system performance measures and report the summary descriptive statistics in the following sections.

6.2.1 Number of Departures

Aircraft automatically departed at scheduled times according to the air traffic scenario unless the participants implemented a departure stop. The mean number of departures that the participants could accept from all airports increased from the Normal condition (122.25) to the Collocated condition (131.38) to the Terminalized condition (133.63). The participants at the EWR and LGA sectors were able to accept 10% more departing aircraft in both the Collocated and Terminalized conditions. Increases in the number of departures accepted were smaller for JFK and TEB, whereas HPN and MMU remained stable across conditions (see Figure 60).

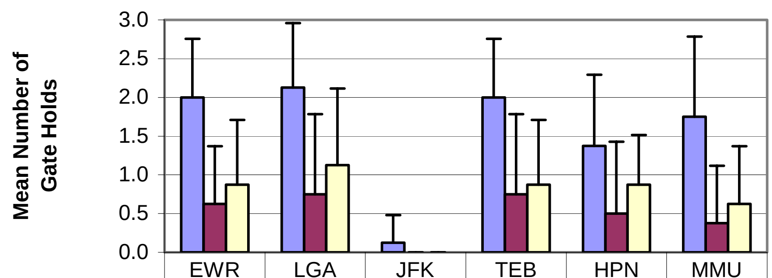


	EWR	LGA	JFK	TEB	HPN	MMU
Normal	40.50	37.63	14.75	15.00	9.50	4.88
Collocated	44.63	41.38	15.13	16.13	9.38	4.75
Terminalized	45.63	42.13	15.13	16.50	9.50	4.75

Figure 60. Mean number of departures by Airport and Condition.

6.2.2 Number of Departure Stops

The terminal participants were able to stop scheduled departures from any airport at their discretion. On average, the participants were able to reduce the number of departure stops by about half at all airports during the Collocated and Terminalized conditions (see Figure 61). The number of departure stops was lowest in the Collocated condition for all airports except JFK, which had no departure stops in either the Collocated or Terminalized conditions. There was a lot of variability between the participants in the number of departure stops observed.

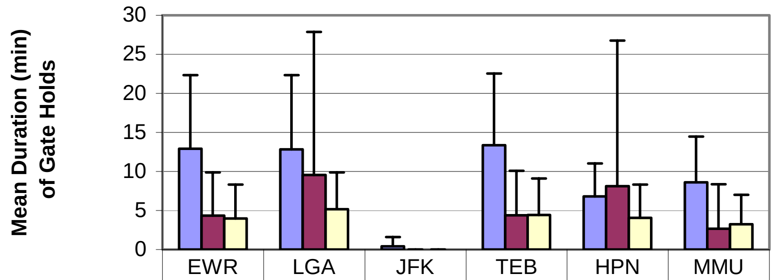


	EWR	LGA	JFK	TEB	HPN	MMU
Normal	2.00	2.13	0.13	2.00	1.38	1.75
Collocated	0.63	0.75	0.00	0.75	0.50	0.38
Terminalized	0.88	1.13	0.00	0.88	0.88	0.63

Figure 61. Mean number of departure stops by Airport and Condition.

6.2.3 Duration of Departure Stops

We calculated the duration of departure stops by measuring the time from the beginning of a departure stop to the end of a departure stop. Overall, the participants greatly reduced the mean duration of departure stops in the Collocated and Terminalized conditions compared to the Normal condition (see Figure 62). At HPN, however, the mean duration of departure stops was longest in the Collocated condition.



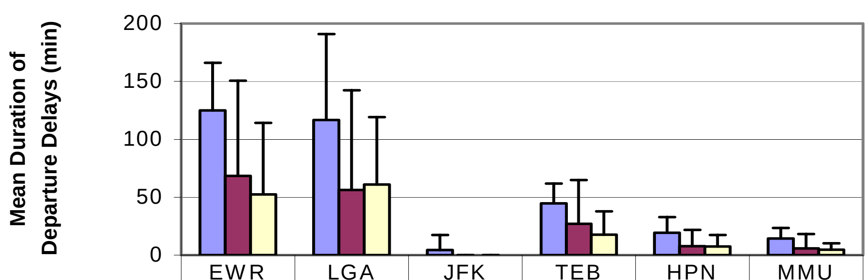
	EWR	LGA	JFK	TEB	HPN	MMU
Normal	12.92	12.84	0.41	13.35	6.78	8.59
Collocated	4.35	9.53	0.00	4.40	8.13	2.67
Terminalized	3.97	5.18	0.00	4.43	4.06	3.24

Airport

Figure 62. Mean duration (min) of departure stops by Airport and Condition.

6.2.4 Duration of Departure Delays

We calculated the duration of departure delays by subtracting each aircraft's scheduled departure time from the actual departure time and then summing across all aircraft departing from an airport. The participants reduced the mean cumulative duration of departure delays by almost half or more at all airports during the Collocated and Terminalized conditions (see Figure 63).



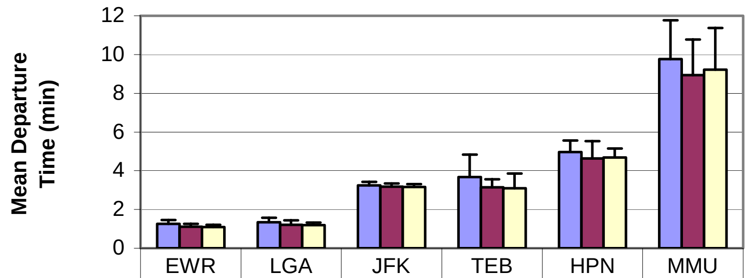
	EWR	LGA	JFK	TEB	HPN	MMU
Normal	125.01	116.62	4.54	44.69	19.18	14.41
Collocated	68.41	56.22	0.00	27.12	7.68	5.72
Terminalized	52.51	61.05	0.00	17.63	7.55	4.64

Airport

Figure 63. Mean cumulative duration (min) of departure delays by Airport and Condition.

6.2.5 Departure Intervals

We defined a departure interval as the time between departures for a given airport. Although the departure intervals (i.e., time between departures) did not differ much between conditions, they were generally longest in the Normal condition and shortest in the Terminalized condition, except for HPN and MMU (see Figure 64).



	EWR	LGA	JFK	TEB	HPN	MMU
Normal	1.25	1.35	3.24	3.67	4.97	9.76
Collocated	1.11	1.20	3.18	3.15	4.64	8.95
Terminalized	1.10	1.19	3.16	3.10	4.68	9.22

Airport

Figure 64. Mean time (min) between departures by Airport and Condition.

6.2.6 Number of Aircraft Handled

The mean number of aircraft handled was highest in the Terminalized condition and lowest in the Normal condition. The participants, across all sectors, handled an average of 4.75 more aircraft per 50-min scenario in the Terminalized condition compared to the Normal condition (see Figure 65). They handled an intermediate number in the Collocated condition.

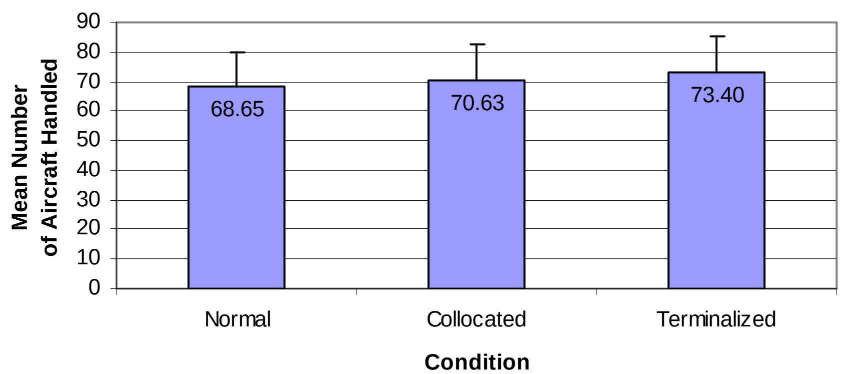
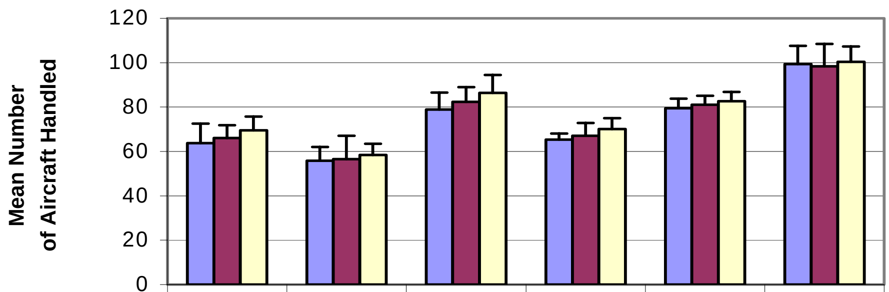


Figure 65. Mean number of aircraft handled collapsed across all Sector Positions (EWR, LGA, LIBW, 39 and 55) by Condition.

When we examined the mean number of aircraft handled by each sector, the participants handled the fewest number of aircraft in the Normal condition, and the participants handled the most aircraft in the Terminalized condition (see Figure 66), except for the Ghost sector. The number of aircraft handled in the Collocated condition was always intermediate between the Normal and Terminalized conditions, except for the Ghost sector.



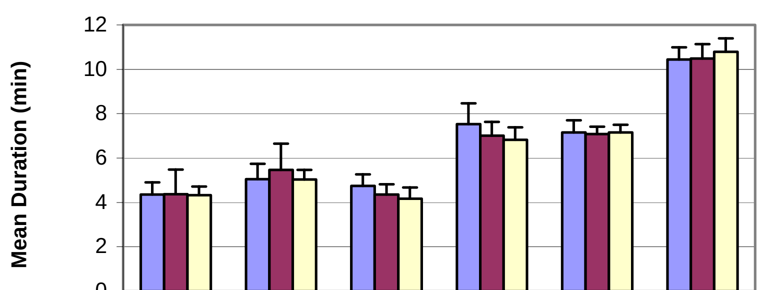
	EWR	LGA	LIBW	39	55	Ghost
Normal	63.75	55.75	78.88	65.38	79.50	99.38
Collocated	66.00	56.50	82.38	67.13	81.13	98.38
Terminalized	69.50	58.38	86.38	70.13	82.63	100.38

Sector Position

Figure 66. Mean number of aircraft handled by Sector Position and Condition.

6.2.7 Duration of Aircraft Handled

We calculated the duration that a participant handled an aircraft by measuring the time each aircraft was on each sector's radio frequency. Figure 67 shows the mean duration that the participants handled each aircraft. There were only very small differences between conditions in most sectors, except for LIBW and 39, which received the most benefit from both collocation and the use of terminal separation standards.



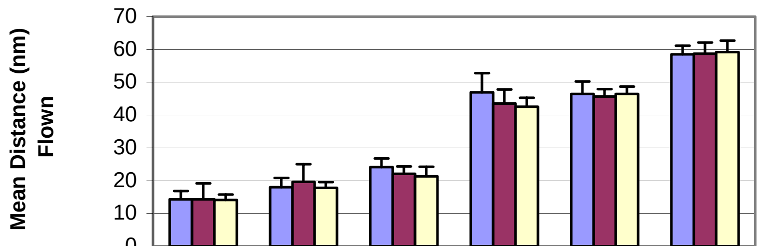
	EWR	LGA	LIBW	39	55	Ghost
Normal	4.35	5.05	4.75	7.53	7.15	10.44
Collocated	4.38	5.47	4.36	7.01	7.08	10.49
Terminalized	4.33	5.03	4.17	6.82	7.15	10.79

Sector Position

Figure 67. Mean duration (min) on the radio frequency per aircraft by Sector Position and Condition.

6.2.8 Distance Flown

The distance-flown variable measured how far the aircraft flew during each simulation run. We calculated the mean distance flown per aircraft to examine the overall efficiency in each experimental condition. We calculated this measure for each sector and condition by dividing the total distance flown in each sector by the number of aircraft handled in the corresponding sector. Although the savings were relatively small (see Figure 68), the Terminalized condition resulted in the shortest distance flown for all of the participant-controlled sectors except for Sector 55. For Sector 55, the shortest distance flown occurred in the Collocated condition. The sector positions LIBW and 39 showed the greatest benefit from both collocation and the use of terminal separation standards. Compared to the Normal condition, they had a reduction of approximately 3 and 4 miles, respectively, in the Terminalized condition. There were intermediate reductions of approximately 2 and 3 miles, respectively, in the Collocated condition.



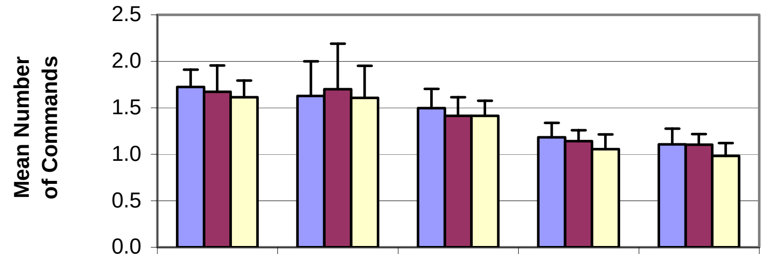
	EWR	LGA	LIBW	39	55	Ghost
Normal	14.31	17.98	24.18	46.88	46.43	58.49
Collocated	14.28	19.68	22.10	43.53	45.68	58.67
Terminalized	14.08	17.78	21.34	42.57	46.47	59.21

Figure 68. Mean distance (nm) flown per aircraft by Condition and Sector.

6.2.9 Number of Altitude Commands

On average, the participants gave about 483 altitude commands in each condition. The terminal sectors gave more altitude commands (ranging from about 92-120) than did the en route sectors (ranging from about 76-86).

Overall, the participants tended to decrease the number of altitude commands per aircraft by about 8-10% in the Terminalized conditions compared to the Normal condition except for the LGA sector (see Figure 69).



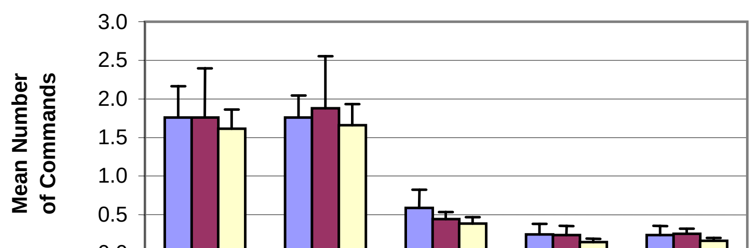
	EWR	LGA	LIBW	39	55
Normal	1.72	1.63	1.50	1.18	1.11
Collocated	1.67	1.70	1.41	1.14	1.10
Terminalized	1.62	1.61	1.41	1.06	0.98

Figure 69. Mean number of altitude commands per aircraft by Sector Position and Condition.

6.2.10 Number of Heading Commands

On average, the participants gave about 281 heading commands during each scenario. The participants gave an average of 53 heading commands per sector in the Terminalized condition compared to 58 in the Normal and Collocated conditions. The number of heading commands contributed more to the taskload of the EWR and LGA participants (about 113 and 99, respectively) than to the en route participants (about 13 and 17, respectively). This result is not surprising because the en route participants could fly the aircraft straight through their sectors, vectoring only for traffic, whereas the terminal participants had to vector the aircraft for departure.

Figure 70 shows the mean number of heading commands per aircraft by Sector Position and Condition. Although the differences are small, the data trend suggests that the participants in all sectors gave fewer heading commands per aircraft in the Terminalized condition compared to the Normal and Collocated conditions. The sector positions LIBW and 39 showed the largest percentage reductions.



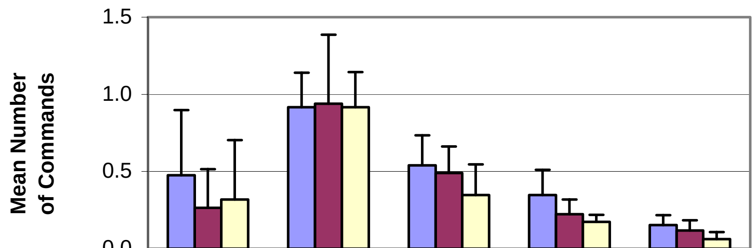
	EWR	LGA	LIBW	39	55
Normal	1.76	1.76	0.59	0.24	0.24
Collocated	1.76	1.88	0.44	0.23	0.25
Terminalized	1.61	1.66	0.38	0.14	0.16

Figure 70. Mean number of heading commands per aircraft by Sector Position and Condition.

6.2.11 Number of Speed Commands

On average, the participants gave about 137 speed commands per scenario. The participants gave about 25 speed commands per sector in the Collocated and Terminalized conditions compared to about 32 speed commands per sector in the Normal condition. The mean cumulative number of speed commands shows that speed commands generated taskload primarily at the terminal sectors, especially LGA (about 51) and LIBW (about 37). The Collocated and Terminalized conditions reduced taskload in terms of the total number of speed commands that the participants had to issue in the EWR, LIBW, 39, and 55 sectors.

Figure 71 shows the mean number of speed commands issued per aircraft. The same pattern of results was obtained as for the total number of speed commands issued. The Collocated condition reduced the number of speed commands that the participants gave per aircraft at the EWR sector position; however, the Terminalized condition resulted in more consistent and greater benefits, especially for the LIBW, 39, and 55 sector positions.



	EWR	LGA	LIBW	39	55
Normal	0.48	0.92	0.54	0.35	0.15
Collocated	0.26	0.94	0.49	0.22	0.12
Terminalized	0.32	0.92	0.35	0.17	0.06

Sector Position

Figure 71. Mean number of speed commands per aircraft by Sector Position and Condition.

Overall, in terms of clearances issued, the Terminalized condition provided the greatest benefit. The Terminalized condition did result in some slight increases in taskload for some terminal sectors; however, it resulted in decreases in taskload for the en route sectors. Across all sector positions, the number of total commands either stayed the same or decreased in the Terminalized condition compared to the Normal condition despite the fact that the participants handled more aircraft. The participants were also able to reduce the number of commands given per aircraft in the Terminalized condition compared to the Normal condition. The Collocated condition also resulted in slight decreases in the number of commands given per aircraft compared to the Normal condition, but not to the same extent and with the same consistency as in the Terminalized condition.

6.3 Safety Measures

We examined how often pairs of aircraft lost standard separation and the number of operational errors to assess the potential air traffic risk in each condition. We did not account for aircraft that were on diverging headings or aircraft that may have been using visual separation procedures. We also examined the frequency of wake turbulence violations. We used the same separation criterion in this experiment as we used in the previous experiment (see Section 3.3 Safety Measures).

6.3.1 Loss of Separation

We conducted separate analyses for the en route and terminal loss of separation data. For the en route loss of separation data set, we conducted a 2 (Sector) X 3 (Condition) repeated measures ANOVA. There was a significant main effect of Condition, $F(2, 14) = 10.22$. There were significantly fewer losses of separation occurrences in the Terminalized condition compared to the Normal and Collocated conditions, HSD(14) = 2.60. Figure 72 shows the en route loss of separation data.

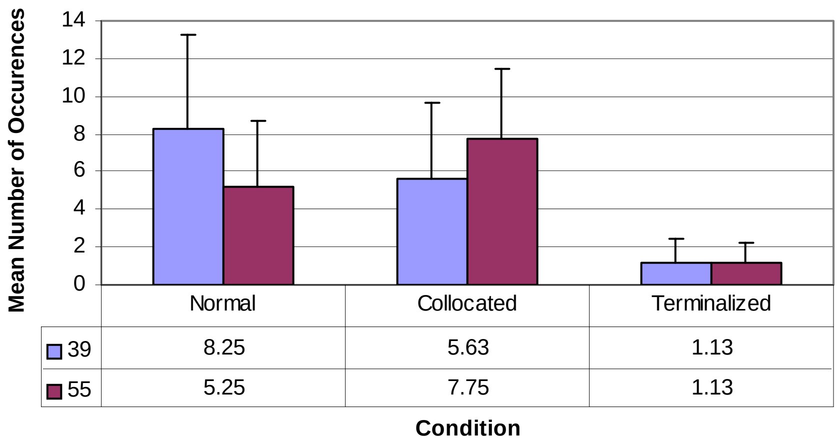


Figure 72. Mean number of occurrences of en route loss of separation by Sector and Condition.

For the terminal loss of separation data, we conducted a 3 (Sector) X 3 (Condition) mixed-effects ANOVA with sector as the between-subjects variable and Condition as the within-subjects variable. None of the effects were significant. Figure 73 shows the terminal loss of separation data.

We also examined the number of en route operational errors using the same criteria as in the arrival experiment. However, we did not examine each operational error in depth to eliminate extraneous causes because this process would have been too time consuming. Rather, we assume that any extraneous factors that may have affected the data were equally and randomly distributed across the three conditions. We analyzed the en route operational error data by conducting a 2 (Sector) X 3 (Condition) repeated measures ANOVA. None of the effects were significant. Overall, operational errors occurred equally often in each sector and condition (see Figure 74).

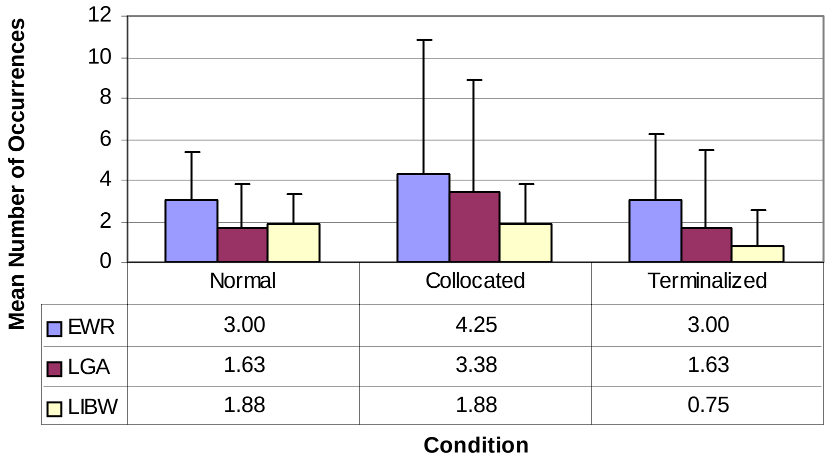


Figure 73. Mean number of occurrences of terminal loss of separation by Sector and Condition.

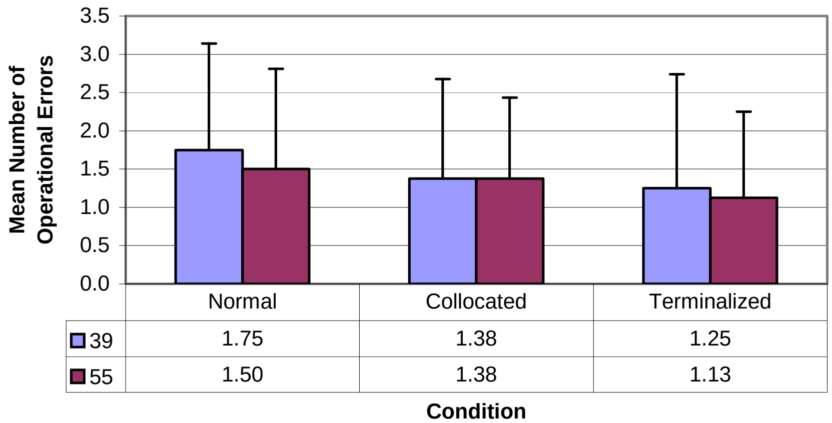


Figure 74. Mean number of en route operational errors by Sector and Condition.

6.3.2 Wake Turbulence

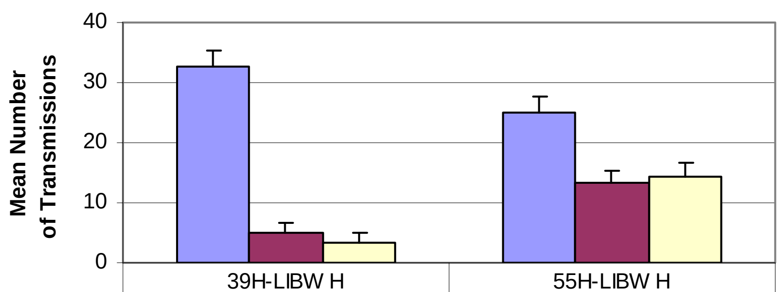
We calculated wake turbulence violations based on FAA Order 7110.65N (FAA 2002d) with some exceptions. We only examined aircraft pairs that were above 1000 ft msl. We assumed that aircraft below that altitude were the responsibility of the tower controller. The number of en route wake turbulence violations ranged from 3.50 in the Collocated condition to 5.06 in the Terminalized condition. The number of terminal wake turbulence violations ranged from 4.71 in the Normal condition to 6.71 in the Collocated condition. The statistical analyses showed that wake turbulence violations occurred equally often in all sectors and conditions for both the en route and terminal participants.

6.4 Communication Measures

For communications behavior, we examined the PTT and the FTF communication data for both the en route and terminal participants. We examined each of these data sets separately for each domain. For the PTT data, we examined the mean number and duration of ground-ground and ground-air transmissions within each domain.

6.4.1 En Route Push-to-Talk Communications

We analyzed the en route transmissions using a 2 (Sector Position: 39H, 55H) X 3 (Condition: Normal, Collocated, Terminalized) repeated measures ANOVA. The ground-ground en route communications included all landline transmissions that went from sector position 39H to LIBW H and from sector position 55H to LIBW H. There was a significant Sector Position X Condition interaction, $F(2, 14) = 6.72$ (see Figure 75). The interaction indicates that sector position 39H reduced the number of their ground-ground transmissions during the Collocated and Terminalized conditions more than did sector position 55H. For the ground-ground transmissions from sector position 39H to LIBW H, there were significantly fewer transmissions in the Collocated and Terminalized conditions compared to the Normal condition, HSD(14) = 11.89. These results indicate that collocation was beneficial in reducing landline communication between LIBW H and both en route sectors, but it was significantly more beneficial when the sectors were side by side.



	39H-LIBW H	55H-LIBW H
Normal	32.75	25.13
Collocated	5.13	13.50
Terminalized	3.50	14.25

Sector Position

Figure 75. Mean cumulative number of en route ground-ground landline transmissions by Sector Position and Condition.

For the mean duration of ground-ground landline transmissions, there were significant main effects of Sector Position, $F(1, 7) = 7.13$, and Condition, $F(2, 14) = 6.56$ (see Figure 76). The participants made longer transmissions when they worked at Sector Position 55H compared to when they worked at Sector Position 39H. The post hoc test on the main effect of Condition was marginal, indicating that landline transmissions were shorter in the Collocated condition compared to the Normal condition. Similar to the number of ground-ground landline transmissions, the duration results show the benefit of being side by side so that information can be shared visually and through gestures.

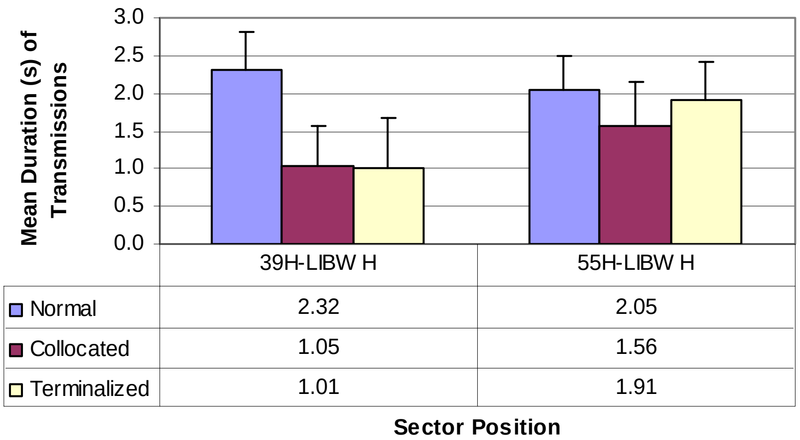


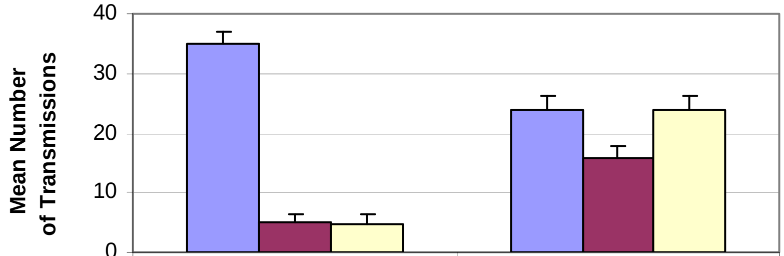
Figure 76. Mean duration (s) of en route ground-ground landline transmissions by Sector Position and Condition.

The en route ground-air transmissions were those transmissions that went from a participant to a simulation pilot that was flying an aircraft inside the ZNY or N90 airspace. The participants made a similar number of ground-air transmission when they worked at Sector Position 39 (Mean = 260.25, *SD* = 4.45) compared to when they worked at Sector Position 55 (Mean = 266.50, *SD* = 4.61). The number of ground-air transmissions also remained stable across the conditions. The mean number of ground-air transmissions made during each 50 min scenario ranged from 253 to 274. There were no significant effects for the duration of en route ground-air radio transmissions either. On average, each en route ground-air transmission took about 3.4 s.

6.4.2 Terminal Push-to-Talk Communications

The terminal ground-ground communications included all landline transmissions that went from Sector Position LIBW H to 39H and from Sector Position LIBW H to 55H. We analyzed the terminal ground-ground transmissions using a 2 (Sector Position: LIBW-39H, LIBW-55H) X 3 (Condition: Normal, Collocated, Terminalized) repeated measures ANOVA. There was a significant Sector Position X Condition interaction for the mean number of ground-ground landline transmissions, $F(2, 14) = 11.06$ (see Figure 77). The post hoc test of the interaction was significant, HSD(14) = 15.34. For ground-ground transmissions from LIBW H to Sector 39H, the participants made fewer transmissions in the Collocated and Terminalized conditions compared to the Normal condition. The same reduction did not occur for ground-ground transmissions from LIBW H to Sector 55H. Furthermore, in both the Collocated and Terminalized conditions, the participants made fewer transmissions from LIBW H to Sector 39H compared to transmissions made from LIBW H to Sector 55H. The results are very similar to the en route data, but show that LIBW H had to contact the physically more distant Sector 55H more often to coordinate.

There were no significant effects for the mean duration of terminal ground-ground landline transmissions. On average, each terminal ground-ground landline transmission took about 1.7 s, which remained stable across the conditions.



	LIBW H-39H	LIBW H-55H
Normal	34.88	24.00
Collocated	4.88	15.75
Terminalized	4.75	24.00

Sector Position

Figure 77. Mean cumulative number of terminal ground-ground landline transmissions by Sector Position and Condition.

The terminal ground-air transmissions were those transmissions that went from a terminal participant to a simulation pilot. We analyzed the terminal ground-air transmissions using a 3 between (Sector Position: EWR, LGA, LIBW) X 3 within (Condition: Normal, Collocated, Terminalized) mixed-effects ANOVA. For the mean number of terminal ground-air transmissions, there was a significant main effect of Sector Position, $F(2, 21) = 29.75$, and the post hoc test was significant, HSD(21) = 31.34. The participants at the LIBW (Mean = 337.58, *SD* = 8.78) sector position made significantly more ground-air transmissions than either EWR (Mean = 281.42, *SD* = 8.78) or LGA (Mean = 242.33, *SD* = 8.78). The participants at the EWR sector position also made significantly more ground-air transmissions than LGA. The mean number of ground-air transmissions was relatively stable across conditions (Mean = 287.11, *SD* = 52.06). These results reflect the mean number of aircraft handled in each sector.

There were no significant effects for the mean duration of terminal ground-air transmissions. On average, each terminal ground-air transmission took about 3.4 s. This duration remained stable across sector positions and conditions.

For both the terminal and en route sector positions, the Collocated and Terminalized conditions resulted in fewer ground-ground landline transmissions, especially those between Sector Positions 39H and LIBW. We expected this result because the wall separating the en route and terminal sectors was not present during these conditions, and the participants could engage in FTF communication. The en route participants also reduced the duration of ground-ground landline transmissions during the Collocated condition in comparison to the Normal condition.

6.4.3 Face-to-Face Communications

The participants may have compensated for the reduction in the number of landline communications during the Collocated and Terminalized conditions, especially those communications between Sector Positions LIBW and 39, by engaging in more FTF communication. FTF communication between the terminal and en route sectors was possible only in the Collocated and Terminalized conditions because experimenters removed the wall separating terminal and en route sector positions during these two conditions. In this section, we examine the communication data collected using the CSS.

The participants did take advantage of the opportunity for FTF communication. On average, there were 40.1 verbal communications in the Collocated condition and 45.0 verbal communications in the Terminalized condition. In addition to exchanging verbal information, they also used the opportunity to acquire information by looking (glancing) at one another's radar display. Table 15 shows the mean number of basic communication behaviors by Condition and Type of behavior. We used the same coding scheme for type of communication as in the arrival experiment, except that the communication involved Sector Positions LIBW, 39, and 55. Most of the FTF communications were verbal, followed by glances. The observer recorded relatively few behaviors in the remaining categories.

Table 15. Mean Number of Basic Communication Behaviors by Condition and Type

	Glance	Verbal	Non-verbal	Other	Non-ATC	Could not code
Collocated	5.25	40.13	1.00	0.00	0.00	0.00
Terminalized	5.13	45.00	3.50	0.00	0.00	0.00

The observer classified and recorded the participants' verbal communications as in the arrival experiment. The participants exchanged information about route most frequently, followed by information about traffic and altitude (see Table 16). There was also some exchange of information regarding Aircraft Speeds and Flow Messages. There were no communications that the observer would have classified as a Point Out.

Table 16. Mean Number of Verbal Communication Behaviors by Condition and Type

	Approval	Handoff	Point Out	Traffic	Altitude	Route	Speed	Flow Msg	FPS	A/C ID
Collocated	0.25	0.75	0.00	8.75	7.25	12.88	5.00	4.63	0.13	0.25
Terminalized	1.38	5.63	0.00	8.38	8.38	16.25	1.50	3.38	0.13	0.00

We can further examine the participants' FTF communication behavior by understanding who was directing the communication to whom. Table 17 shows the mean number of basic communication behaviors by Condition, Sector Position, and Type. We coded whether communications were going from Sector Positions LIBW or LIBW H to 39 or 55 or from Sector Position 39 or 55 to LIBW/LIBW H. We did not distinguish if communications that were going from Sector Position 39 or 55 to LIBW or LIBW H because either participant at the LIBW radar or handoff positions were just as likely to receive the communication, and we did not want to make any assumptions about who actually received the communication. The observer recorded glances that the participants exchanged between Sector Positions LIBW and 39, but did not observe any glances between LIBW and 55. The LIBW sector position was located next to Sector Position 39, so most of the communication and coordination took place between these two sector positions. The participants at Sector Positions LIBW and 55 did not glance at each others radar scope, but there was a good deal of verbal information exchanged between these two sector positions.

Table 17. Mean Number of Basic Communication Behaviors by Condition, Sector Position, and Type

		Glance	Verbal	Non-verbal
Collocated	LIBW←39	3.00	8.50	0.25
	LIBW←39H	0.00	7.88	0.13
	LIBW←55	0.00	1.00	0.13
	LIBW←55H	0.00	3.63	0.00
	LIBW→39	1.88	8.75	0.25
	LIBW→55	0.00	0.88	0.00
	LIBWH→39	0.38	5.75	0.13
	LIBWH→55	0.00	3.75	0.13
Terminalized	LIBW←39	2.38	9.25	0.88
	LIBW←39H	0.13	8.63	0.50
	LIBW←55	0.00	1.25	0.00
	LIBW←55H	0.25	4.25	0.13
	LIBW→39	1.63	9.63	1.13
	LIBW→55	0.00	0.88	0.25
	LIBWH→39	0.63	6.88	0.50
	LIBWH→55	0.13	4.25	0.13

6.5 ATC Observer Rating Form

6.5.1 En Route Observer Rating Form

Nearly all of the SME performance ratings on the en route ORF were in the high range on the 8-point scales. We analyzed each item of the en route ORF using a 4 (Sector Position: 39, 39H, 55, 55H) X 3 (Condition: Normal, Collocated, Terminalized) repeated measures ANOVA. We obtained significant main effects of Sector Position on 7 of the 26 scales (overall safe and efficient traffic flow, maintaining awareness of aircraft positions, detecting pilot deviations, correcting their own errors in a timely manner, overall attention and SA, overall prioritizing, and marking flight strips while performing other tasks), but none of the post hoc tests were significant. In all cases, the SMEs rated Sector Positions 39 and 39H approximately one point higher than Sector Positions 55 and 55H. All the ratings except marking flight strips while performing other tasks indicated high performance in all positions (39/39H ranged from 6.29 – 6.83 and 55/55H ranged from 5.28 – 6.13). Marking flight strips was rated as 5.90 and 6.28 for 39 and 39H, and as 4.24 and 4.05 for 55 and 55H, respectively. The SME may have given higher ratings to Sector 39 because that sector may not have been as complex as Sector 55, and thus performance was actually better. The SME reported that when LIBW became busy, complexity was reduced at Sector 39. Sector 55 was also working arrival traffic that had to be descended into a lower sector (Washington/BWI complex), which may have made their task more difficult.

There was a significant main effect of Condition for the participants' rated performance on handling control tasks for several aircraft, $F(2, 12) = 3.90$. The SMEs used this item to assess the participants' ability to shift control tasks between aircraft and their ability to perform timely communications while time-sharing with other actions. The data trend indicates that the participants received the highest rating in the Terminalized condition and the lowest rating in the Normal condition (see Figure 78).

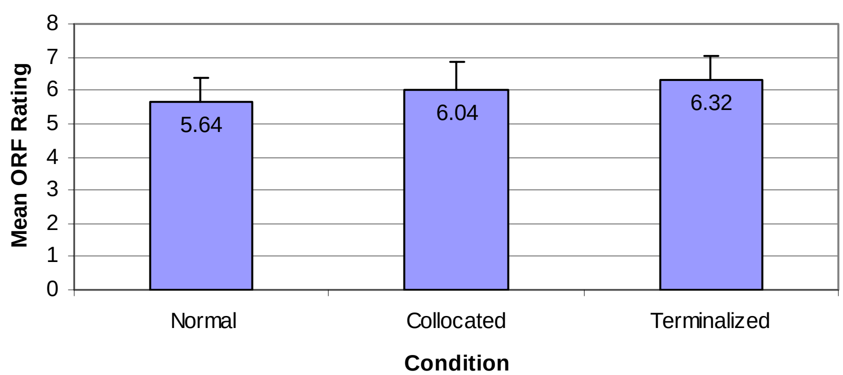


Figure 78. Mean en route ORF ratings for handling control tasks for several aircraft by Condition.

There was a significant main effect of Condition, $F(2, 12) = 8.95$, for the participants rated performance for marking FPSs while performing other tasks (see Figure 79). The post hoc test was not significant. However, the data trend indicates that the participants received the highest mean ratings during the Terminalized condition.

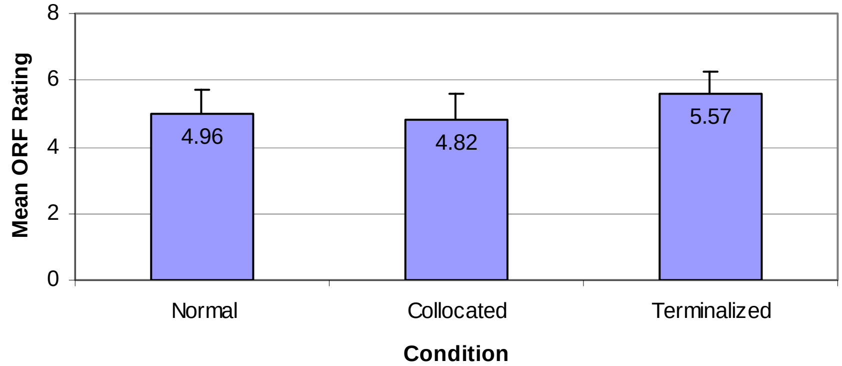


Figure 79. Mean en route ORF ratings for marking flight progress strips while performing other tasks by Condition.

6.5.2 Terminal Observer Rating Form

Nearly all the SME ratings on the terminal ORF were also in the high performance range. We analyzed each item of the terminal ORF using a 2 (Sector Position: R-side, Handoff) X 3 (Condition: Normal, Collocated, Terminalized) repeated measures ANOVA. We conducted separate analyses for each terminal sector (i.e., EWR, LGA, LIBW). We report the statistically significant effects in the following paragraphs.

For EWR, there was a significant main effect of Condition for the participants rated performance on preplanning control actions, $F(2, 14) = 8.71$ (see Figure 80). The SME used this item to rate the participant's ability to scan the adjacent sectors to plan for future and conflicting traffic and the participant's ability to study pending FPSs in the bay. The post hoc test indicated that the participants received significantly higher mean ratings in the Normal condition compared to the Collocated condition, HSD(14) = 0.51. The ratings in the Terminalized condition were not significantly different from either the Normal or Collocated conditions.

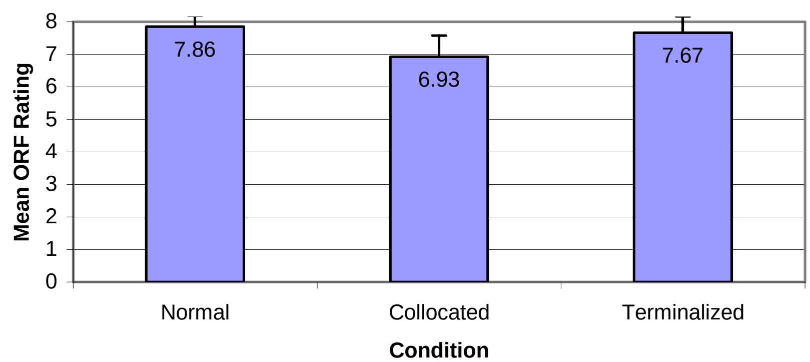


Figure 80. Mean EWR ORF ratings for preplanning control actions by Condition.

There was a significant main effect of Condition in EWR for the SME ratings of the participants performance on providing coordination, $F(2, 14) = 3.78$ (see Figure 81). The SME used this item of the ORF to assess the participants' ability to provide effective and timely coordination and to use proper point-out procedures. The data trend suggests that the participants were best able to provide coordination in the Terminalized condition and least able to provide coordination in the Collocated condition. However, the ratings were high for all of the conditions.

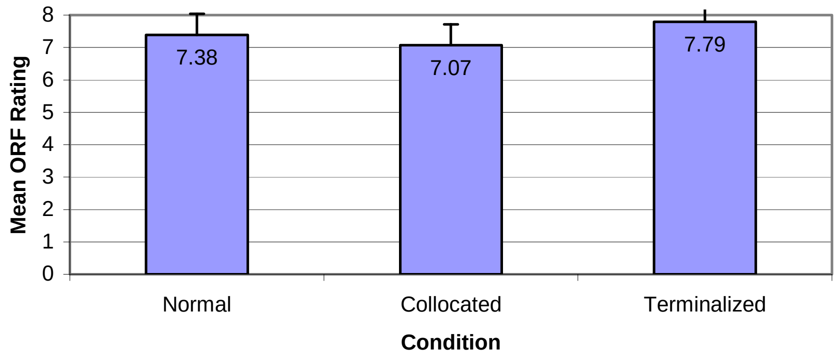


Figure 81. Mean EWR ORF ratings for providing coordination by Condition.

For the LGA sector, there was a significant main effect of Condition for the participants' rated ability to maintain separation and resolve conflicts, $F(2, 14) = 5.48$. This item assessed the participants' ability to use control instructions that maintain appropriate aircraft and airspace separation, detect and resolve impending conflicts early, and recognize the need for speed restrictions and wake turbulence separation. The marginal post hoc test indicates that the mean ratings were higher in the Terminalized conditions than in the Collocated condition. The SME rated performance as being very high in all conditions though (see Figure 82).

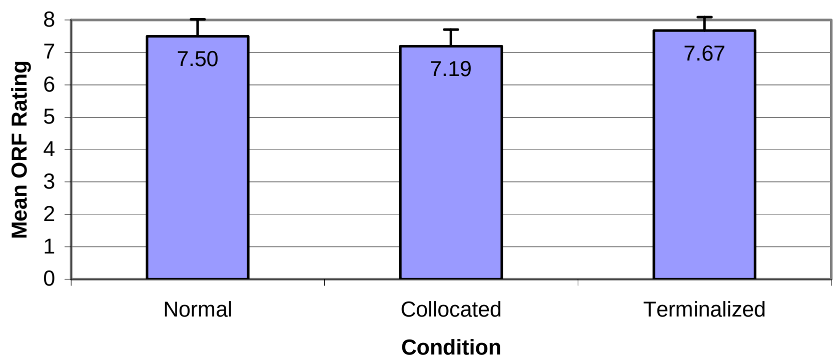


Figure 82. Mean LGA ORF ratings for maintaining separation and resolving potential conflicts by Condition.

There were no significant differences for the SME ratings of the participants' performance at LIBW. The participants at LIBW received very high ORF ratings in all conditions and at both the R-side and handoff positions.

6.6 Workload Measures

The participants provided subjective measures of workload by responding to the WAK and NASA-TLX.

6.6.1 Workload Assessment Keypad Ratings

We analyzed the mean WAK ratings separately for en route and terminal participants. If a participant did not respond to a WAK prompt, we treated those data as missing for the same reasons as in the arrival experiment (see Section 3.6.1).

6.6.1.1 En Route Workload Assessment Keypad Ratings

Overall, the en route participants rated their workload as being low to moderate. For the analysis of the en route WAK ratings, we used a 3 (Condition) X 4 (Sector Position) repeated measures ANOVA. There was a marginal effect of Condition (see Figure 83). Although the differences were not statistically significant, the data trend shows that the participants rated workload the lowest in the Terminalized condition. The participants reported the highest levels of workload in the Normal condition and intermediate levels of workload in the Collocated condition.

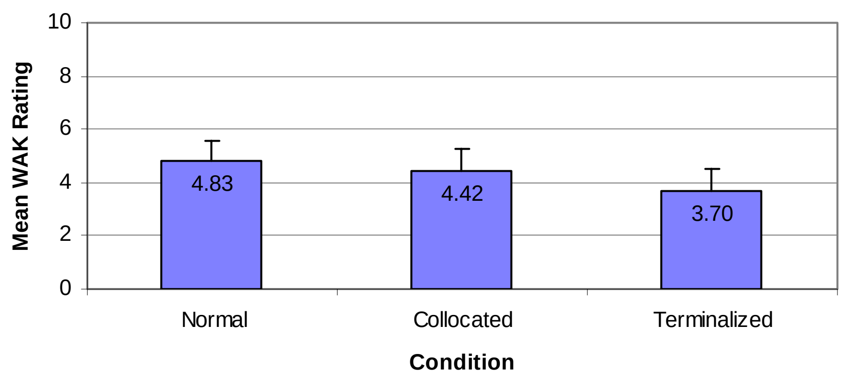


Figure 83. Mean en route WAK ratings by Condition.

6.6.1.2 Terminal Workload Assessment Keypad Ratings

For the analysis of the terminal WAK ratings, we conducted a separate 3 (Condition) X 2 (Sector Position) repeated measures ANOVA for each Sector (EWR, LGA, LIBW).

The participants' WAK ratings at the EWR sector were significantly higher at the R-side position (Mean = 3.05, SD = 0.82) than at the handoff position (Mean = 2.23, SD = 0.74), $F(1, 7) = 18.56$. Overall, the EWR participants perceived workload to be low across all conditions and sector positions.

The participants' WAK ratings at the LGA sector were also significantly higher at the R-side position (Mean = 2.45, SD = 0.51) compared to the handoff position (Mean = 2.12, SD = 0.62), $F(1, 7) = 6.71$. The participants at the LGA sector also perceived workload to be low in all conditions and sector positions.

There was a marginal main effect of Condition for the participants' WAK ratings at the LIBW sector (see Figure 84). The data trend shows the highest WAK ratings occurred in the Terminalized condition, followed by the Collocated condition and then the Normal condition. Overall, the LIBW participants perceived workload as being moderate to slightly high in all conditions and sector positions.

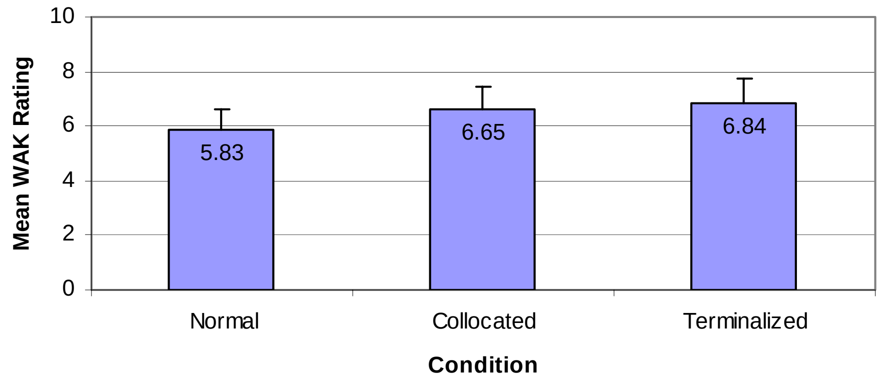


Figure 84. Mean LIBW WAK ratings by Condition.

6.6.2 NASA Task Load Index

We analyzed the en route and terminal TLX ratings separately. Items on the TLX asked participants about their mental demand, physical demand, temporal demand, effort, frustration, and performance.

6.6.2.1 En Route Task Load Index

For the en route participants' ratings, we analyzed each item of the TLX separately using a 3 (Condition: Normal, Collocated, Terminalized) X 4 (Sector Position: 39, 39H 55, 55H) repeated measures ANOVA.

There was a significant main effect of Condition for the participants' ratings of mental demand, $F(2, 14) = 6.22$ (see Figure 85). The data trend suggests that the effect was due to the higher ratings of mental demand in the Normal condition compared to the Terminalized condition.

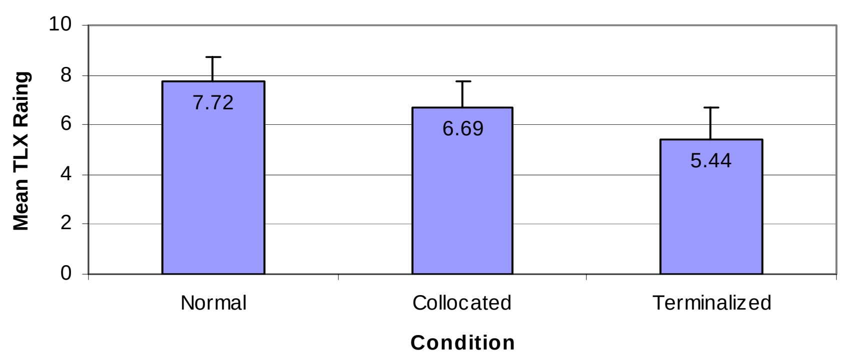
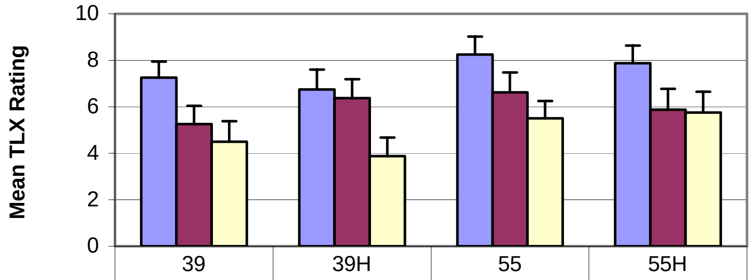


Figure 85. Mean en route TLX ratings for mental demand by Condition.

There were significant main effects of Sector Position and Condition for the participants' ratings of temporal demand, $F(3, 21) = 3.68$ and $F(2, 14) = 7.74$, respectively (see Figure 86). The data trend suggests that the participants perceived temporal demand to be higher in the Normal condition than in the Terminalized condition. Temporal demand also appeared to be higher at Sector Position 55 compared to Sector Position 39.



	39	39H	55	55H
Normal	7.25	6.75	8.25	7.88
Collocated	5.25	6.38	6.63	5.88
Terminalized	4.50	3.88	5.50	5.75

Figure 86. Mean en route TLX ratings for temporal demand by Sector Position and Condition.

There was a significant main effect of Condition for the participants' ratings of effort, $F(2, 14) = 5.30$ (see Figure 87). The data trend indicates that the participants thought that they had to expend more effort in the Normal condition compared to the Terminalized condition.

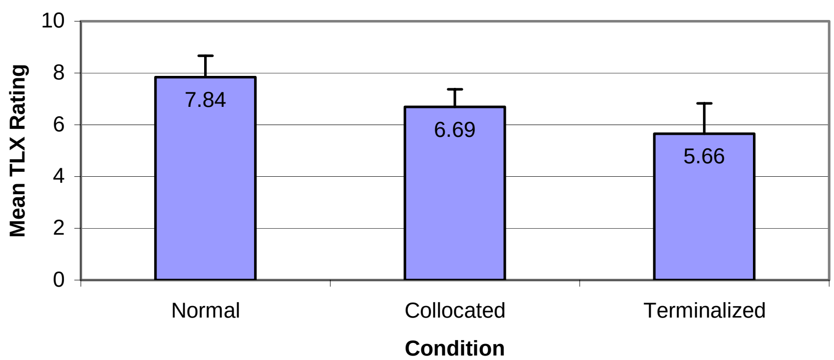


Figure 87. Mean en route TLX ratings for effort by Condition.

There was also a significant main effect of Condition for the participants' ratings of frustration, $F(2, 14) = 11.06$ (see Figure 88). The participants reported that they experienced more frustration in the Normal condition compared to the Terminalized condition.

The ratings for physical demand were moderate to slightly high and ranged from 7.3 (1.00) in the Normal condition to 6.5 (1.02) in the Collocated condition to 5.5 (1.18) in the Terminalized condition. There was a marginal main effect of Condition for these ratings. The ratings of

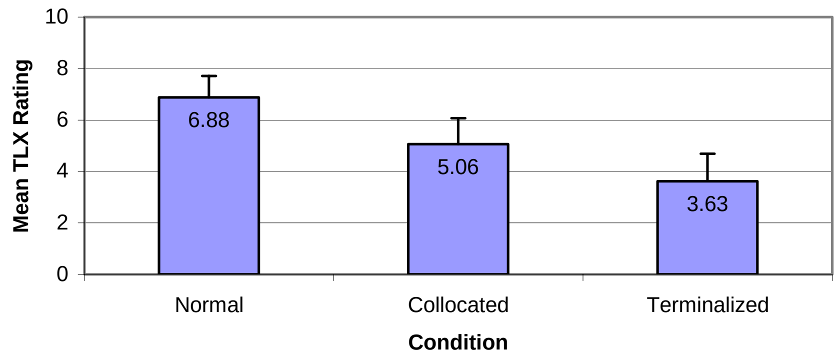


Figure 88. Mean en route TLX ratings for frustration by Condition.

en route performance were high in all conditions. The participants' ratings of their performance ranged from 6.8 (1.07) in the Normal condition to 8.3 (0.93) in the Terminalized condition.

Overall, the en route participants favored the Terminalized condition over the Normal condition. The participants reported that the Terminalized condition produced less mental demand, temporal demand, effort, and frustration compared to the Normal condition. The participants gave intermediate ratings for the Collocated condition.

6.6.2.2 Terminal Task Load Index

For the analysis of the terminal TLX ratings, we conducted a separate 3 (Condition) X 2 (Sector Position) repeated measures ANOVA for each Sector (EWR, LGA, LIBW).

In the EWR sector, the TLX ratings were in the low to moderate range on the 10-point scales, except for performance, which was high. There was a significant main effect of Sector Position for the EWR participants' ratings of mental demand, $F(1, 7) = 12.85$. The participants perceived mental demand to be higher for the R-side position (Mean = 4.96, $SD = 0.82$) than the handoff position (Mean = 3.08, $SD = 0.92$).

There was a significant main effect of Sector Position for the EWR participants' ratings of physical demand, $F(1, 7) = 9.99$. The participants rated physical demand as being higher when they worked the R-side position (Mean = 3.79, $SD = 0.88$) compared to the handoff position (Mean = 2.67, $SD = 0.96$).

There was also a main effect of Sector Position for the EWR participants' ratings of temporal demand, $F(1, 7) = 16.53$. The participants reported experiencing greater temporal demand while working the R-side position (Mean = 4.04, $SD = 0.87$) compared to the handoff position (Mean = 2.75, $SD = 0.97$).

This was a main effect of Sector Position for the EWR participants' ratings of effort, $F(1, 7) = 14.25$. The participants rated their effort as being higher while working the R-side position (Mean = 5.42, $SD = 0.84$) compared to the handoff position (Mean = 3.25, $SD = 0.86$).

Finally, there were main effects of Sector Position and Condition for the EWR participants' ratings of frustration, $F(1, 7) = 11.89$ and $F(2, 14) = 3.95$, respectively. The data trend shows that the participants' rated their level of frustration as being higher while working the R-side position compared to the handoff position (see Figure 89). Furthermore, they perceived greater frustration during the Normal condition than during the Terminalized condition.

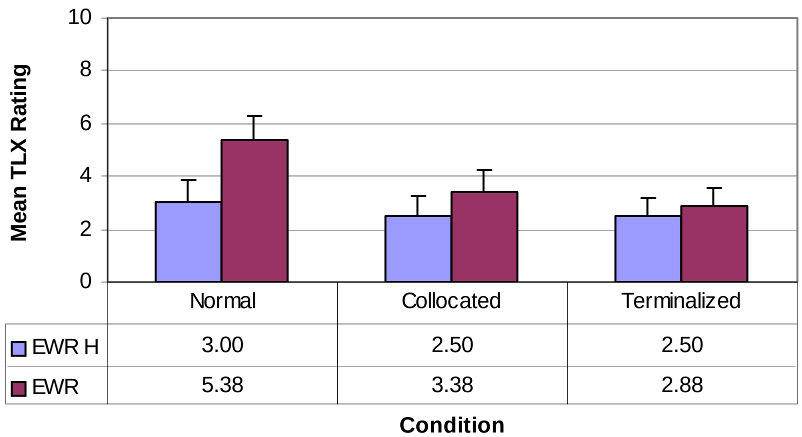


Figure 89. Mean EWR TLX ratings for frustration by Sector Position and Condition.

In the LIBW sector, the TLX ratings were generally high, and there were only two significant effects. There was a significant main effect of Condition for the participants' ratings of mental demand, $F(2, 14) = 5.68$ (see Figure 90). The marginal post hoc test suggests that mental demand was greater during the Normal condition compared to the Collocated condition.

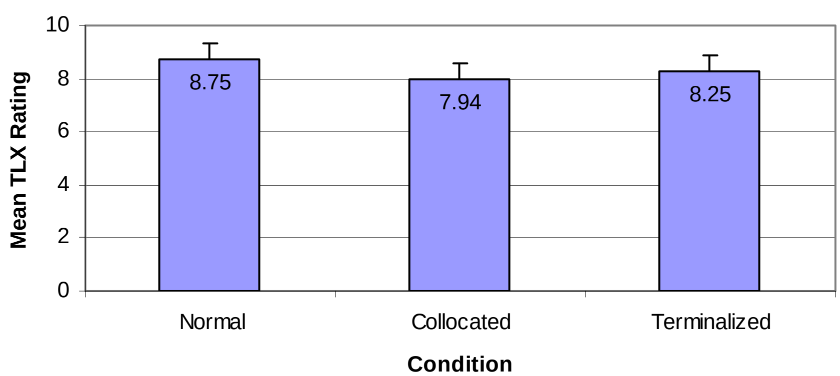


Figure 90. Mean LIBW TLX ratings for mental demand by Condition.

There was also a significant main effect of sector position for the LIBW participants' ratings of temporal demand, $F(1, 7) = 9.69$. When they worked at the R-side position (Mean = 8.04, $SD = 0.82$), the LIBW participants reported greater temporal demand than when they worked at the handoff position (Mean = 7.04, $SD = 0.87$).

The terminal participants' TLX ratings did not clearly favor one condition over another. The EWR participants experienced greater frustration in the Normal condition compared to the Terminalized condition, and the LIBW participants thought they had higher mental demand in the Normal condition compared to the Collocated condition. The TLX ratings do not show many advantages for the Collocated or Terminalized conditions, but they do not show any detriments either.

6.7 Questionnaires

6.7.1 Post-Scenario Questionnaire

This section presents the data collected from the participants immediately after they completed each 50-min scenario.

6.7.1.1 En Route Post-Scenario Questionnaire

Except for workload due to ground-ground communications, the en route PSQ ratings were in the moderate to high range. To determine if there were differential effects, we analyzed each item of the en route PSQ separately using a 3 (Experimental Condition: Normal, Collocated, Terminalized) X 4 (Sector Position: 39, 39H, 55, 55H) repeated measures ANOVA. Table 18 shows all of the means and *SD*s for the en route PSQ items by Sector Position and Condition except for the two questions about the reduced lateral separation standards, which only applied to the Terminalized condition.

Table 18. Means and Standard Deviations for the En Route PSQ Items
by Sector Position and Condition

PSQ Item	Sector Position	Normal	Collocated	Terminalized
1. Rate your overall level of ATC performance during this scenario.	39	7.00 (2.13)	7.62 (2.07)	8.62 (1.30)
	39H	6.75 (1.67)	7.13 (1.64)	8.63 (1.19)
	55	7.25 (1.83)	7.63 (2.45)	8.25 (0.71)
	55H	6.25 (2.12)	7.25 (2.19)	7.88 (2.47)
2. Rate your ability to move aircraft through the sector during this scenario.	39	6.75 (1.98)	7.38 (1.69)	9.12 (0.99)
	39H	6.62 (2.45)	6.88 (2.70)	8.88 (1.13)
	55	6.13 (1.25)	7.50 (2.73)	9.38 (0.52)
	55H	6.13 (2.17)	7.75 (1.67)	8.63 (1.51)
3. Rate your overall level of SA during this scenario.	39	7.00 (2.14)	8.00 (1.41)	8.62 (1.06)
	39H	7.12 (1.36)	7.00 (2.51)	8.38 (1.30)
	55	7.13 (1.46)	7.13 (2.47)	8.13 (0.99)
	55H	5.88 (2.03)	6.75 (2.25)	8.13 (1.64)
4. Rate your SA for current aircraft locations during this scenario.	39	6.63 (1.69)	7.50 (2.45)	8.63 (1.30)
	39H	6.75 (1.16)	6.63 (2.20)	8.50 (1.77)
	55	6.25 (2.12)	6.63 (2.45)	7.87 (0.64)
	55H	5.13 (2.23)	6.88 (2.75)	7.88 (1.55)
5. Rate your SA for projected aircraft locations during this scenario.	39	6.25 (2.12)	8.25 (1.39)	8.63 (1.19)
	39H	6.57 (1.90)	6.50 (2.78)	8.00 (1.60)
	55	5.75 (2.43)	6.63 (2.50)	7.63 (1.77)
	55H	5.25 (2.19)	7.00 (2.00)	7.75 (1.75)
6. Rate your SA for potential aircraft loss-of-separation during this scenario.	39	5.88 (1.88)	7.88 (1.13)	8.63 (1.77)
	39H	7.00 (1.51)	7.25 (2.43)	7.62 (2.50)
	55	6.63 (2.33)	7.00 (2.51)	8.25 (1.58)
	55H	5.50 (2.45)	7.13 (2.17)	8.25 (2.05)
7. Rate your workload due to air-to-ground communications during this scenario.	39	6.88 (2.03)	6.00 (2.27)	5.38 (1.69)
	39H	4.38 (2.33)	4.63 (2.88)	3.25 (1.83)
	55	8.50 (1.31)	6.75 (2.71)	6.75 (2.49)
	55H	3.50 (1.60)	4.50 (1.85)	3.75 (1.91)
8. Rate your workload due to ground-to-ground communications during this scenario.	39	3.88 (1.89)	3.00 (1.31)	2.75 (1.75)
	39H	4.25 (1.67)	4.13 (1.89)	3.87 (2.42)
	55	3.13 (2.30)	2.88 (2.42)	2.25 (1.16)
	55H	5.38 (2.00)	4.00 (2.00)	3.88 (2.10)
9. Rate the performance of the simulation pilots in terms of their responding to your control instructions and providing readbacks.	39	8.13 (1.89)	8.25 (1.49)	8.25 (1.75)
	39H	8.38 (1.41)	8.75 (1.28)	9.13 (1.46)
	55	8.75 (1.28)	8.75 (1.49)	8.87 (1.25)
	55H	8.38 (1.41)	8.13 (2.10)	8.88 (1.13)
10. Rate the difficulty of this scenario.	39	7.38 (1.60)	7.50 (1.77)	6.88 (1.55)
	39H	7.50 (1.85)	7.75 (1.16)	7.13 (2.64)
	55	7.88 (2.23)	8.00 (1.60)	7.50 (2.33)
	55H	7.63 (1.51)	7.75 (1.83)	7.50 (1.69)

There was a significant main effect of Condition for the participants' ratings of their ability to move aircraft through the sector, $F(2, 14) = 13.93$ (see Figure 91). They indicated that they were better able to move aircraft through the sector during the Terminalized condition than during the Normal condition.

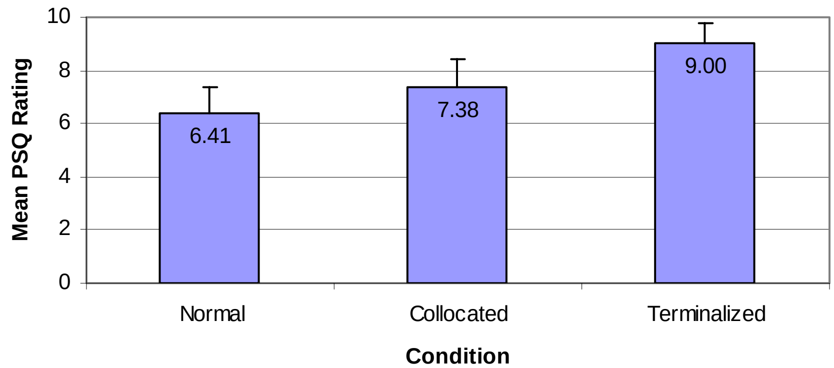


Figure 91. Mean en route PSQ ratings for ability to move aircraft through the sector by Condition.

There was a significant main effect of Condition for the participants' ratings of their overall level of SA, $F(2, 14) = 5.41$ (see Figure 92). They indicated that they had better overall SA in the Terminalized condition compared to the Normal condition.

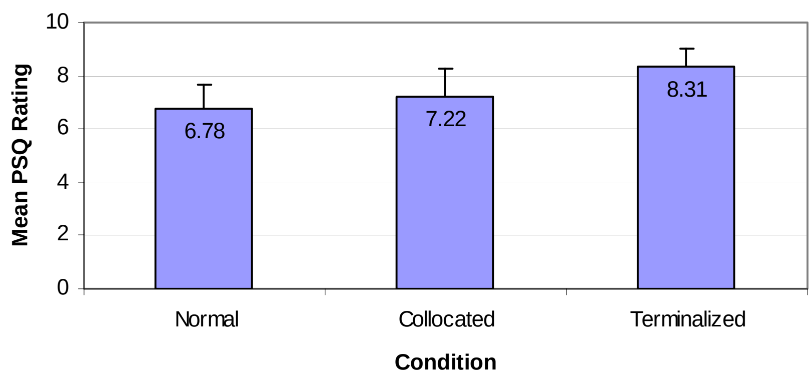


Figure 92. Mean en route PSQ ratings for overall level of SA by Condition.

There was a significant main effect of Condition for the participants' ratings of their SA for current aircraft locations, $F(2, 14) = 6.61$ (see Figure 93). They reported having better SA for current aircraft locations in the Terminalized condition compared to the Normal condition.

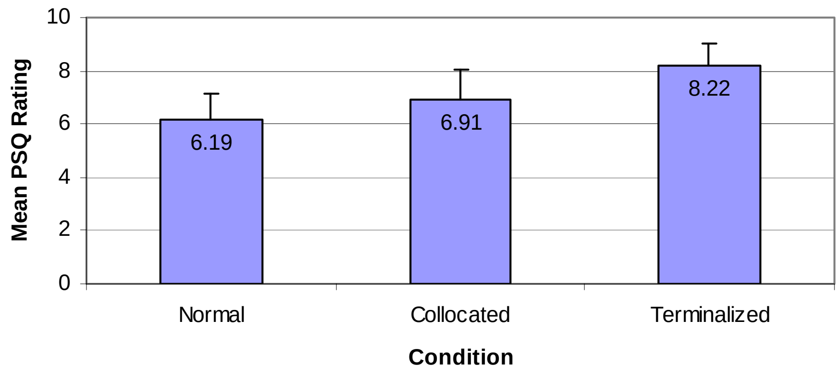


Figure 93. Mean en route PSQ ratings for SA for current aircraft locations by Condition.

There was also a significant main effect of Condition for the participants' ratings of their SA for projected aircraft locations, F(2, 14) = 5.86 (see Figure 94). Again, they indicated that they had better SA for projected aircraft locations in the Terminalized condition compared to the Normal condition.

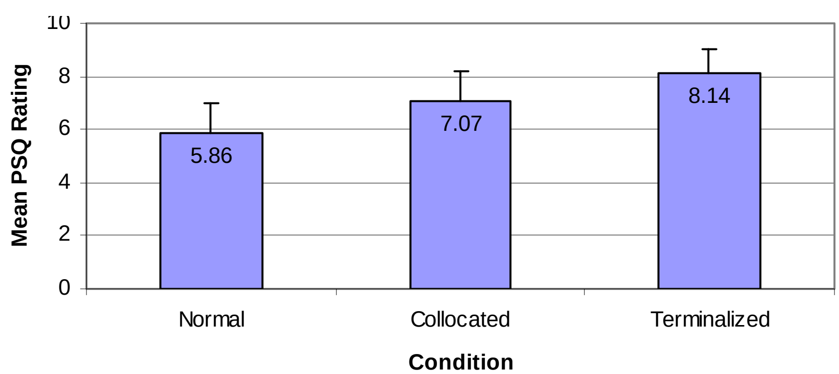


Figure 94. Mean en route PSQ ratings for SA for projected aircraft locations by Condition.

The participants reported that they had significantly better SA for potential aircraft loss of separation in the Terminalized condition compared to the Normal condition, $F(2, 14) = 10.31$ (see Figure 95).

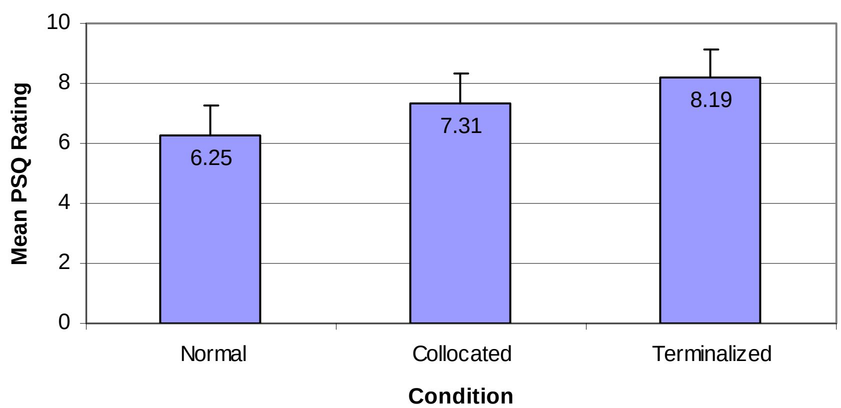


Figure 95. Mean en route PSQ ratings for SA for potential loss of aircraft separation by Condition.

The participants reported significantly different levels of workload at the various sector positions due to ground-air communications, $F(3, 21) = 8.51$ (see Figure 96). The marginal post hoc test indicated that the participants thought they had higher workload at Sector Position 55 compared to 55H. This result was not surprising because the R-side was primarily responsible for communicating with aircraft.

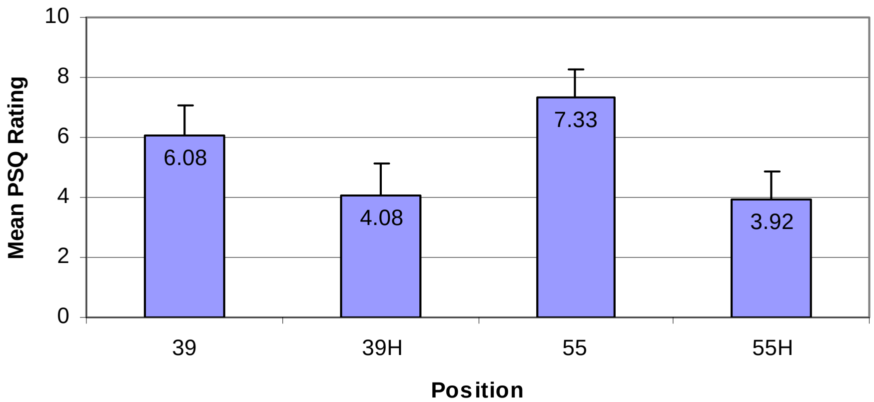


Figure 96. Mean en route PSQ ratings for workload due to ground-air communication by Sector Position.

There were two additional questions on the PSQ that only applied to the Terminalized condition. All of the participants indicated that the reduced lateral separation standard had a positive impact on their ability to control aircraft (mean rating of 8.4 across all sector positions). They also indicated they were able to adapt to those lateral separation standards (mean rating of 8.9 across all sector positions). There was also little variability in either set of ratings (standard deviations ranged from 0.47 to 0.80). Compared to the Normal condition, the en route participants preferred the Terminalized condition for the ability to move aircraft through the sector, overall SA, SA for current aircraft locations, SA for projected aircraft locations, and for SA for potential loss of aircraft separation. The mean ratings for the Collocated condition were intermediate between the Normal and Terminalized conditions.

6.7.1.2 Terminal Post-Scenario Questionnaire

Table 19 shows the means and *SD*s for the terminal PSQ items by Sector Position and Condition except for the two questions about the reduced lateral separation standards, which only applied to the Terminalized Condition. Except for the questions about communications workload and scenario difficulty, the ratings were generally in the high to very high range. To determine if there were any differential effects for each terminal Sector (EWR, LGA, LIBW), we analyzed each item of the terminal PSQ separately using a 3 (Experimental Condition: Normal, Collocated, Terminalized) X 2 (Sector Position: R-side, handoff) repeated measures ANOVA.

Table 19. Means and Standard Deviations for the Terminal PSQ Items
by Sector Position and Condition

PSQ Item	Sector Position	Normal	Collocated	Terminalized
1. Rate your overall level of ATC performance during this scenario.	EWR	8.13 (1.46)	8.13 (1.96)	8.75 (1.39)
	EWR H	7.88 (2.64)	8.13 (1.81)	8.00 (2.00)
	LGA	8.75 (1.91)	8.75 (1.83)	8.63 (1.30)
	LGA H	8.00 (2.14)	7.50 (3.16)	8.13 (2.17)
	LIBW	7.50 (2.33)	8.50 (1.07)	8.00 (1.20)
	LIBW H	6.88 (1.55)	7.38 (1.69)	7.25 (1.49)
2. Rate your ability to move aircraft through the sector during this scenario.	EWR	8.13 (1.73)	8.63 (2.00)	9.38 (0.74)
	EWR H	6.63 (2.97)	8.38 (1.69)	7.88 (2.03)
	LGA	8.63 (1.77)	8.88 (1.89)	8.88 (1.36)
	LGA H	7.50 (2.14)	7.50 (3.16)	7.88 (2.30)
	LIBW	7.63 (1.69)	8.13 (1.81)	8.50 (1.07)
	LIBW H	6.38 (2.33)	6.63 (1.60)	7.63 (0.92)
3. Rate your overall level of SA during this scenario.	EWR	8.00 (1.60)	8.50 (2.00)	9.25 (1.04)
	EWR H	7.63 (2.92)	7.75 (2.12)	8.13 (2.10)
	LGA	8.86 (1.86)	8.38 (1.77)	8.63 (1.30)
	LGA H	8.00 (2.27)	7.38 (3.07)	7.75 (3.06)
	LIBW	8.13 (1.13)	8.13 (1.89)	8.50 (0.76)
	LIBW H	6.25 (2.19)	7.25 (2.31)	7.00 (1.41)
4. Rate your SA for current aircraft locations during this scenario.	EWR	8.13 (1.46)	8.88 (2.03)	9.38 (0.74)
	EWR H	7.88 (2.23)	8.13 (2.17)	8.25 (2.31)
	LGA	8.63 (1.85)	8.25 (2.05)	8.88 (1.36)
	LGA H	8.00 (2.07)	7.25 (3.11)	7.88 (2.80)
	LIBW	7.38 (1.92)	8.00 (1.85)	7.88 (0.99)
	LIBW H	6.50 (2.07)	7.38 (1.77)	6.75 (1.49)
5. Rate your SA for projected aircraft locations during this scenario.	EWR	8.38 (1.30)	8.75 (1.98)	9.38 (0.74)
	EWR H	7.50 (2.00)	8.25 (2.19)	8.13 (2.36)
	LGA	8.38 (1.77)	8.25 (2.19)	8.63 (1.41)
	LGA H	7.75 (2.05)	7.50 (3.12)	8.13 (2.80)
	LIBW	7.25 (1.16)	7.75 (1.83)	7.88 (0.64)
	LIBW H	6.50 (1.93)	7.00 (1.93)	7.00 (1.31)
6. Rate your SA for potential aircraft loss-of-separation during this scenario.	EWR	9.13 (1.13)	8.75 (2.05)	9.00 (1.41)
	EWR H	7.75 (1.91)	8.13 (1.96)	8.38 (1.77)
	LGA	9.00 (1.41)	9.13 (1.81)	9.38 (0.74)
	LGA H	8.75 (1.83)	8.00 (3.12)	8.63 (2.77)
	LIBW	6.75 (2.12)	7.75 (1.91)	8.50 (0.76)
	LIBW H	5.88 (2.10)	7.63 (1.69)	7.50 (1.60)
7. Rate your workload due to air-to-ground communications during this scenario.	EWR	6.25 (2.71)	5.13 (3.04)	6.13 (3.31)
	EWR H	2.63 (1.92)	2.13 (0.83)	2.50 (2.14)
	LGA	5.25 (2.92)	5.00 (3.07)	5.13 (2.17)
	LGA H	3.43 (2.64)	2.63 (1.77)	2.25 (1.16)
	LIBW	7.63 (1.77)	7.63 (2.00)	7.88 (2.59)
	LIBW H	5.75 (2.25)	6.00 (1.31)	5.38 (2.00)

Table 19. Means and Standard Deviations for the Terminal PSQ Items
by Sector Position and Condition (Cont.)

PSQ Item	Sector Position	Normal	Collocated	Terminalized
8. Rate your workload due to ground-to-ground communications during this scenario.	EWR	6.00 (2.62)	3.88 (2.64)	5.63 (3.50)
	EWR H	4.13 (2.53)	3.00 (0.93)	4.13 (2.95)
	LGA	2.00 (1.31)	2.13 (1.55)	2.38 (1.85)
	LGA H	5.38 (2.00)	4.25 (1.58)	3.63 (2.00)
	LIBW	5.13 (1.96)	5.25 (2.76)	6.13 (2.85)
	LIBW H	6.00 (2.62)	5.13 (1.13)	5.63 (2.50)
9. Rate the performance of the simulation pilots in terms of their responding to your control instructions and providing readbacks.	EWR	7.63 (2.20)	8.13 (1.73)	8.25 (1.75)
	EWR H	7.75 (2.96)	7.38 (2.72)	8.38 (1.30)
	LGA	8.25 (2.38)	8.00 (2.20)	8.75 (0.88)
	LGA H	8.25 (2.60)	8.75 (1.16)	8.75 (0.89)
	LIBW	7.43 (2.07)	7.50 (2.07)	7.25 (2.12)
	LIBW H	7.88 (1.89)	8.13 (1.37)	7.25 (1.98)
10. Rate the difficulty of this scenario.	EWR	6.00 (1.69)	5.25 (2.49)	6.13 (1.73)
	EWR H	5.13 (2.42)	5.13 (2.10)	5.38 (1.85)
	LGA	4.50 (2.00)	4.75 (2.12)	5.13 (1.81)
	LGA H	4.38 (2.33)	4.00 (2.00)	4.25 (2.19)
	LIBW	9.00 0.93)	8.75 (1.04)	8.88 (0.99)
	LIBW H	8.63 (1.06)	8.75 (1.04)	8.38 (1.98)

The EWR, LGA, and LIBW participants reported significantly greater workload at the R-side position compared to the handoff position for workload due to ground-air communications, $F(1, 7) = 11.50$, $F(1, 6) = 13.03$, and $F(1, 7) = 17.09$, respectively. As with Sector Positions 55 and 55H, this was an obvious consequence of the R-side performing all of the ground-air communications.

The LGA participants reported significantly greater workload at the handoff position (Mean = 4.42, $SD = 0.93$) compared to the R-side position (Mean = 2.17, $SD = 0.94$) for workload due to ground-ground communications, $F(1, 7) = 9.40$, but this was also a consequence of the division of responsibilities.

The LIBW and LIBW H participants reported significant differences between conditions in their SA for potential loss of separation across conditions, $F(2, 14) = 4.24$ (see Figure 97). The data trend suggests that the participants perceived that they had better SA in the Terminalized condition compared to the Normal condition.

The PSQ proved sensitive for detecting some differences between the terminal sector positions, but most were related to the manual division of combined responsibilities between the R-side and handoff positions. The only difference between conditions occurred when the LIBW and LIBW H participants rated their SA as being higher for the potential loss of separation in the Terminalized condition compared to the Normal condition.

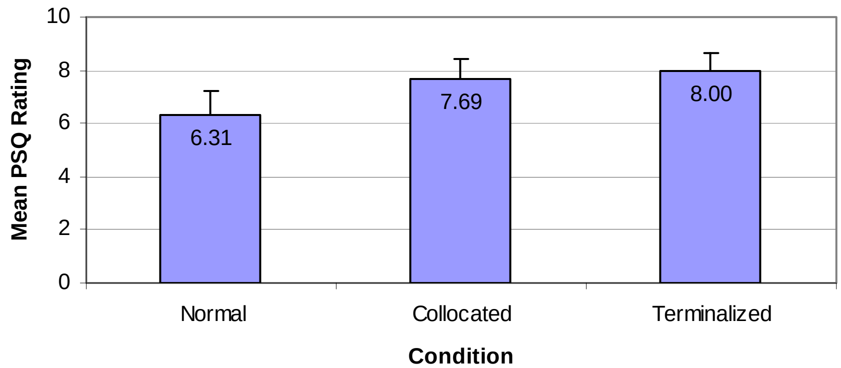


Figure 97. Mean LIBW/LIBW H PSQ ratings for SA for potential loss of separation by Condition.

The terminal participants indicated the reduced lateral separation standards had a positive effect on their ability to control traffic, but to a lesser degree than the en route participant ratings. The terminal participant ratings ranged from 6.3 to 8.0. They also were able to adapt to the use of the reduced separation standards (means ranged from 6.7 to 8.7), but again to a lesser degree than in en route sectors. There was also slightly greater variability in the ratings, with *SD*s ranging from 0.72 to 1.24. These results indicate the terminal participants received some benefits in terms of traffic flow even though the actual change in separation standards only occurred in the en route sectors.

6.7.2 Post-Experiment Questionnaire

All participants completed the PEQ at the end of the experiment. The participants rated Items 1 and 3 of the PEQ using a 10-point scale where a rating of 1 indicated *no change at all* and a 10 indicated *a great deal of change*. The participants rated Items 2 and 4 using a 9-point scale where a rating of 1 indicated a *negative effect*, a rating of 9 indicated a *positive effect*, and a rating of 5 indicated *no effect*. Table 20 shows the ratings from the en route and terminal participants, respectively.

Table 20. Means and Standard Deviations of En Route and Terminal Participants' PEQ Responses

	En Route – ZNY	Terminal – N90
PEQ Item	**Mean (*SD*)**	**Mean (*SD*)**
1. Did your communication strategies change during the Collocated Condition?	8.00 (0.76)	6.00 (3.11)
2. What effect, if any, did collocation alone have on your control strategies?	7.13 (1.13)	6.38 (1.69)
3. Did your communication strategies change during the Terminalized Condition?	8.75 (0.71)	4.46 (3.34)
4. What effect, if any, did the Terminalized Condition have on your control strategies?	8.88 (0.35)	7.04 (1.89)

The en route participants thought that their communications strategies changed substantially during both the Collocated and Terminalized conditions. They also thought that collocation alone had a moderately positive effect on their control strategies and that the Terminalized condition had a highly positive effect on their control strategies. The terminal participants thought that the communications strategies changed somewhat in both the Collocated and Terminalized conditions. They thought that collocation alone had only a slight positive effect on their control strategies, whereas the Terminalized condition had a moderately positive effect.

7. CONCLUSION – DEPARTURE EXPERIMENT

We found a number of benefits for both the Collocated and Terminalized conditions. Compared to the Normal condition, both the Collocated and Terminalized conditions allowed the participants to handle more aircraft, increase the number of departures they could accept, decrease the time between departures, reduce the number and duration of departure stops, and reduce the duration of departure delays. Overall, the Terminalized condition resulted in the greatest system performance benefits. The Terminalized condition did result in more clearances issued in the terminal sectors, but a concomitant reduction occurred in the en route sectors. The number of clearances given per aircraft was lowest in the Terminalized condition.

The Terminalized condition resulted in significantly fewer losses of separation in both en route sectors compared to both the Normal and Collocated conditions. There were no other significant safety effects.

The participants were able to take advantage of their collocated situation, especially Sector Positions LIBW and 39, during the Collocated and Terminalized conditions by reducing the number of ground-ground landline transmissions. In these conditions, the participants engaged in FTF communication in lieu of the landline. The exchange of information between the terminal and en route participants changed quantitatively and qualitatively during the Collocated and Terminalized conditions. The participants exchanged more information in these conditions not only by engaging in FTF communication, but also by acquiring information from each others' radar display. This expanded awareness allowed the participants to broaden their view of the air traffic situation thereby performing in a more strategic, rather than tactical, manner.

The SME performance ratings were high for nearly all items of the ORF. For the en route participants, some of the ratings were higher for Sector Position 39, but the ratings did not vary significantly across the conditions. For the terminal participants, significant differences between conditions were uncommon. The significant differences that were found in the participants' performance between conditions suggested that the Collocated condition may have posed some problems, at least from the SMEs' perspective. During the Collocated condition, the EWR participants were rated as less able to preplan for control actions and provide coordination, and the LIBW participants were rated as less able to maintain separation and resolve potential conflicts. Although these ratings were significantly lower in the Collocated condition, they were still 7 or above on the 8-point scale, indicating that the participants' performance was very good.

Overall, the en route participants reported moderate workload in the Normal condition, and that it declined in the Collocated condition and declined even more in the Terminalized condition. On the TLX, they rated their mental demand, temporal demand, effort, and frustration significantly lower in the Terminalized condition. The terminal participants in EWR and LGA rated their WAK workload as moderate to low in all conditions and sector positions, whereas the LIBW participants rated their WAK workload as moderate in the Normal condition and somewhat increased in the Collocated and Terminalized conditions. However, the maximum LIBW rating was a 6.84 on a 10-point scale. The EWR participants rated their frustration level as higher in the Normal condition compared to the Terminalized condition. The LIBW participants rated their mental demand as significantly higher in the Normal condition compared to the Collocated condition.

According to the PSQ responses, the en route participants thought that they were best able to move aircraft through their sector during the Terminalized condition. They also rated themselves as having better overall SA, better SA for current and projected aircraft locations, and better SA for potential loss of separation during the Terminalized condition. The LIBW participants also reported better SA for the potential loss of aircraft separation in the Terminalized condition.

Both the terminal and en route participants indicated on the PEQ that their communication and control strategies were positively changed in the Collocated and Terminalized conditions. Overall, the participants thought the Terminalized condition provided the greatest positive change, but more so for the en route sectors than the terminal sectors.

8. GENERAL DISCUSSION AND CONCLUSIONS

The overall results of the arrival and departure experiments are supportive of the NYICC concepts of collocation of terminal and en route sectors and of reducing the lateral separation standards in some current en route sectors. Both experiments showed positive effects on the efficiency of the airspace and a reduction in landline communications while improving SA of the controllers. Many of the positive effects were relatively small on an absolute scale, but they are consistent with the benefits found in the FFPO analyses and are likely to be operationally beneficial. An additional one or two aircraft handed off to the tower or an additional four or five departures accepted per 50 min scenario do not seem like much of a benefit, but in practical terms, it is very valuable because of its cumulative impact on any particular airport and on the overall NAS. Delays in the New York metropolitan area compound across the entire country. Another example is the approximately 2 min reduction per aircraft in inbound holding, but that 2 min represents not only a time saving but also a substantial savings to the airlines in fuel costs. In other cases, such as the reduction in duration of departure stops from nearly 13 min in the Normal condition to approximately 4 or 5 min in the Terminalized condition, the effects were obviously operationally beneficial.

The Terminalized condition, which included both collocation of terminal and en route sectors and the use of terminal separation standards in one or both en route sectors, was consistently the most effective environment. The Collocated condition frequently showed some improvements in efficiency compared to the Normal condition, but generally not to the extent found in the Terminalized condition. On a few measures, performance in the Collocated condition was actually lower than in the Normal condition, but only by a small amount. Collocation definitely improved coordination and SA and reduced landline communications, but terminal separation standards were required to optimize the efficiency of the airspace.

The effects shown in the departure simulation were somewhat stronger than in the arrival simulation, probably because it was a much more complex situation that presented more opportunities for the participants to take advantage of FTF communication and the reduced separation criteria. That is, the arrival simulation was primarily a linear problem with most aircraft transiting from ZOB through 75, then 74, then ARD, then landing at EWR, although there were other aircraft entering the simulation from other sectors and centers. This linearity was due to the design of the simulation and more benefits may occur in the real world. The participants reported that collocation and terminal separation standards improved their ability to move aircraft through the sectors, but the objective indices showed relatively small benefits. The departure simulation had two departure sectors feeding to LIBW, which then fed both Sectors 39 and 55, in addition to aircraft departing from four satellite airports. Here, the objective indices more strongly supported the participant evaluation of the benefits of the two proposed changes. This difference is also consistent with the FFPO findings that in simplified situations, the benefits are not as obvious as in situations that are more complex.

There do not appear to be any major negative effects of collocation or reduced separation standards. There were some increases in taskload when sectors were able to handle more aircraft and increases in perceived workload (especially at LIBW), but the increases were relatively small and at their highest levels were still in the moderate to slightly high range, even though they were handling 130% of busy New York traffic days. In some cases, there were shifts in taskload and workload, such as reductions in altitude and heading clearances issued by the EWR participants, whereas there were increases in the number of those clearances by the ARD participants. However, the shift may simply reflect better utilization of the airspace. Another example was that when landline communications were reduced between ARD H and Sector 74, they increased between ARD H and ZDC. This effect may represent a benefit in terms of better FTF coordination with Sector 74 that enables the ARD H participant to spend more time in landline coordination with the external center. The use of terminal separation standards in the en route sectors reduced the number of losses of separation. There were no significant effects of condition on either operational errors or wake turbulence violations.

These two simulations were extremely complex in the context of maintaining sufficient experimental control so that we can draw valid inferences about the impact of collocation and reduced separation compared to normal circumstances, and whether there were any differential impacts to the terminal or en route environments or the particular positions that were staffed. Even though we had 18 participants in the arrival simulation and 32 in the departure simulation, this represents a relatively small sample size on which to base our conclusions. Fortunately, most of the statistical tests were within-subject comparisons so there was less concern about between-subject sampling. In addition, the potential effects are multidimensional, requiring that we take multiple measures with their concomitant risk of erroneously finding effects that are not valid. Some of the measures could not be tested for statistical significance, so they were not evaluated in terms of a probability of chance occurrence. Nonetheless, the overall pattern of results indicates that collocation of terminal and en route sectors would be operationally beneficial, and that use of terminal separation standards can provide even more benefits, at least under heavy traffic loads in current airspace sectors. Some differences may be due to chance variation rather than a direct result of the conditions, but it is extremely unlikely that the overall pattern of findings were due to chance. Neither the Collocated nor the Terminalized condition appears to cause any negative effects.

This conclusion does not mean that all issues associated with the implementation of these proposed changes have been resolved. The following issues were beyond the scope of the current simulations, but were either raised in the caucus discussions or from our analyses of the simulation data and the FFPO reports. First, what impact will the use of collocation and terminalization of ZNY and N90 have on and from other facilities and operational units, such as the tower, traffic flow management, external centers, and the ATCSCC? For example, could the tower accept the increased landing rate and would ground control be able to feed the increased departure rate? Traffic flow restrictions elsewhere may mitigate some of the positive benefits to New York operations, although improvements in New York may reduce the need for restrictions elsewhere.

Second, what will be the interaction between ongoing sector airspace redesign, the introduction of new tools such as URET and TMA, and collocation and reduced separation? Perhaps they will compound the benefits; perhaps they may minimize them. Third, what equipment will the terminalized en route sector use: terminal or en route? In these simulations, the en route controller used the same equipment, except for the 3-mile J ring and just applied reduced separation standards. This was done because we could not change our equipment between scenarios, because we did not want to make different equipment a factor that could affect the en route controllers performance, and because we did not want to use a between-subjects comparison by having a terminal controller unfamiliar with the airspace run the sector during the terminalized condition. Serious consideration should be given to what the system requirements are for terminalized en route sectors, in terms of both the equipment and functionality to be used and how the controllers should be trained.

Fourth, the layout of the consolidated facility requires careful planning to optimize the benefits of collocation. It was quite apparent in the data that, whereas collocation was of some benefit to all sectors, it was of the most benefit to sectors that were physically side-by-side. Most sectors are contiguous with multiple other sectors, so it will be necessary to determine which sectors most need or can most benefit from the overall operation by being in direct physical collocation.

Finally, we recommend that specific procedures be developed and trained to optimize the value of collocation and reduced separation. In these two simulations, we simply instructed the participants that they could interact directly with the other sectors in the experimental conditions, and that specified en route sectors could use 3-nm separation in the Terminalized condition. Presumably, improved procedures and better training would likely increase the benefits derived from both proposed changes. Although anecdotal, we observed differences between groups in their willingness to take advantage of being collocated and of using the reduced separation standards, and a learning curve for individual groups in doing so.

9. EXTERNAL REVIEW PROCESS

As part of their review process, the program office (AEA-520) submitted this report to researchers at NASA Ames for an external review. The purpose of the external review was to examine the scientific and technical merits of this report. The NASA researchers completed two independent reviews. Appendix K includes both reviews along with the authors' responses. The authors' responses are italicized and embedded within the reviewers' text. We also included the final reviewer summary, which indicated that the reviewers had no fundamental scientific concerns about the simulations, that the report explains how the findings are to be interpreted, and that it contains appropriate caveats about the limitations of the study. They concluded that the changes proposed by the authors were acceptable to publish the report. We have made the changes proposed. The project sponsor, AEA-520, has reviewed the report and the NASA comments and has concluded that the simulations have shown the scientific benefit to the collocation of the facilities and expanded terminalization of the airspace. However, further validation efforts are required.

References

American Psychological Association. (2001). *Publication Manual of the American Psychological AssociationI (Fifth Edition).* Washington, DC: Author.

Beatty, R., Corwin, B., & Wambsganss, M. (1999). Collaborative decision-making: A success story of an airline-FAA partnership. In *Proceedings of the Tenth Aviation Psychology Symposium* (pp. 439-444). Columbus, OH: Ohio State University.

Blakey, M. C. (2003). Remarks prepared for delivery to the Federal Aviation Administration 28th Annual Forecast Conference, Washington, DC, March 18, 2003.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching.* Chicago: Rand McNally.

Converse, S. A., Cannon-Bowers, J. A., & Salas, E. (1991). Team member shared mental models: A theory and some methodological issues. In *Proceedings of the Human Factors Society 35th Annual Meeting* (pp. 1417-1421). Santa Monica, CA: Human Factors and Ergonomics Society.

Federal Aviation Administration. (1999). FFP1 performance metrics: An operational impact evaluation plan. Retrieved September 19, 2002, from http://ffp1.faa.gov/approach/media/FFP1metricsplan_v1_AOZ1.doc

Federal Aviation Administration. (2001). Free Flight Program performance metrics results to date. Retrieved September 19, 2002, from http://ffp1.faa.gov/approach/media/June2001_Report.doc

Federal Aviation Administration. (2002a). *Research and Development Strategy.* Washington, DC: Office of Aviation Research.

Federal Aviation Administration. (2002b). *New York Integrated Control Complex. Concept of Operations.* Eastern Region Air Traffic Division.

Federal Aviation Administration. (2002c). Free Flight Program performance metrics results to date. Retrieved September 19, 2002, from http://ffp1.faa.gov/approach/media/June2002_Report.doc

Federal Aviation Administration. (2002d). *Air Traffic Control* (DOT/FAA/Order 7110.65N). Washington, DC: Federal Aviation Administration.

Garvey, J. F. (2001a). Concerning Air Traffic Control delays and capacity. Statement of the Honorable Jane F. Garvey, Federal Aviation Administration Administrator, before the House Committee on Appropriations, Subcommittee on Transportation and Related Agencies. March 15, 2001.

Garvey, J. F. (2001b). *Concerning Air Traffic Control delays and capacity.* Statement of the Honorable Jane F. Garvey, Federal Aviation Administration Administrator, before the House Committee on Appropriations, Subcommittee on Transportation and Related Agencies. May 3, 2001.

Geisser, S., & Greenhouse, S. W. (1958). An extension of Box's results on the use of the *F* distribution in multivariate analysis. *Annals of Mathematical Statistics, 29,* 885-891.

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.

Hawkins, F. H. (1987). *Human factors in flight.* Brookfield, VT: Gower Publishing Co.

Hays, W. L. (1988). *Statistics.* New York: Holt, Rinehart, and Winston.

Huynh, H., & Feldt, L. S. (1970). Conditions under which mean square ratios in repeated measures designs have exact *F* distributions. *Journal of American Statistical Association, 65,* 1582-1589.

Jensen, R. S., & Biegelski, C. S. (1989). Cockpit resource management. In Jensen, R. S. (Ed.), *Aviation psychology* (pp. 176-209). Brookfield, VT: Gower Publishing Co.

Kirk, R. E. (1982). *Experimental design: Procedures for the behavioral sciences* (2^{nd} ed.). Belmont, CA: Brooks-Cole.

MacDonald, L. (1998). Collaborative decision making in aviation. *Journal of Air Traffic Control*, *40*(3), 12-17.

National Airspace Redesign. (2002). NY/NJ/PHL Airspace Redesign. *Program overview presented to the East Coast Aviation Forum*, November 14, 2002.

Orlady, H. W., & Orlady, L. M. (1999). *Human factors in multi-crew flight operations*. Brookfield, VT: Ashgate Publishing Co.

Peterson, L. M., Bailey, L. L., & Willems, B. F. (2001). *Controller-to-controller communication and coordination taxonomy (C^4T)* (DOT/FAA/AM-01/19). Washington, DC: Office of Aerospace Medicine.

Rognin, L., & Blanquart, J. P. (2000). Human communication, mutual awareness and system dependability. Lessons learnt from Air-Traffic Control field studies. *Reliability Engineering and System Safety, 71*(3), 327-336.

Rouanet, H., & Lepine, D. (1970). Comparisons between treatments in a repeated measurement design: ANOVA and multivariate methods. *British Journal of Mathematics and Statistics for Psychology, 23,* 147-163.

Sollenberger, R., Stein, S., & Gromelski, S. (1997). *The development and evaluation of a behaviorally based rating form for assessing Air Traffic Controller performance.* (DOT/FAA/CT-TN 96/16). Atlantic City International Airport, NJ: Federal Aviation Administration, William J. Hughes Technical Center.

Stein, E. S. (1985). *Air Traffic Controller workload. An examination of workload probe*. (DOT/FAA/CT-TN 84/24). Atlantic City International Airport, NJ: Federal Aviation Administration Technical Center.

Stein, E. S., & Rosenberg, B. L. (1983). *The measurement of pilot workload* (DOT/FAA/EM-81/14). Atlantic City International Airport, NJ: U. S. Department of Transportation, Federal Aviation Administration.

Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics (Second Edition).* New York: Harper Collins Publishers, Inc.

Truitt, T. R., Durso, F. T., Crutchfield, J. M., Moertl, P. M., & Manning, C. A. (2000). Test of an Optional Strip Posting and Marking Procedure. *Air Traffic Control Quarterly*, *8*(2), 131-154.

Warkentin, M. E., Sayeed, L., & Hightower, R. (1997). Virtual teams versus face-to-face teams: An exploratory study of a web-based conference system. *Decision Sciences Journal, 28*, 975-996.

Willems, B., & Heiney, M. (2002). *Decision support automation research in the en route Air Traffic Control environment* (DOT/FAA/CT-TN02/10). Atlantic City International Airport, NJ: FAA William J. Hughes Technical Center.

Willems, B., & Truitt, T. R. (1999). *Implications of reduced involvement in en route Air Traffic Control* (DOT/FAA/CT-TN-99/22). Atlantic City International Airport, NJ: Federal Aviation Administration, William J. Hughes Technical Center.

Acronyms

A/C ID	Aircraft Identification
AEA	Eastern Region
ANOVA	Analysis of Variance
ARD	Yardley/Penns Sector
ARTCC	Air Route Traffic Control Center
ATC	Air Traffic Control
ATCSCC	Air Traffic Control System Command Center
ATM	Air Traffic Management
CPC	Certified Professional Controller
CPI	Co-Principal Investigator
CRD	Computer Readout Display
CSS	Communication Score Sheet
DESIREE	Distributed Environment for Simulation, Rapid Engineering, and Experimentation
D-SIDE	Data Side
DSR	Display System Replacement
ERP	Engineering Research Psychologist
EWR	Newark International Airport
FAA	Federal Aviation Administration
FFPO	Free Flight Program Office
FL	Flight Level
FPS	Flight Progress Strip
FTF	Face-to-Face
HPN	Westchester County Airport
HSD	Honestly Significant Difference
JFK	John F. Kennedy International Airport
LGA	LaGuardia Airport
LIBW	Liberty West Sector
LOA	Letters of Agreement
MANOVA	Multivariate Analysis of Variance
MMU	Morristown Municipal Airport
MSL	Mean Sea Level
N90	New York Terminal Radar Approach Control
NAS	National Airspace System
NASA	National Aeronautics and Space Administration
NATCA	National Air Traffic Controllers Association
NYICC	New York Integrated Control Complex
ORF	Observer Rating Form
PEQ	Post-Experiment Questionnaire
PHL	Philadelphia International Airport
PI	Principal Investigator
PSQ	Post-Scenario Questionnaire
PTT	Push-to-Talk
RA	Research Assistant
RDHFL	Research, Development, and Human Factors Laboratory
R-SIDE	Radar Controller

SA	Situation Awareness
SD	Standard Deviation
SME	Subject Matter Expert
SOP	Standard Operating Procedure
STARS	Standard Terminal Automation Replacement System
TEB	Teterboro Airport
TGF	Target Generation Facility
TLX	Task Load Index
TMA	Traffic Management Advisor
TRACON	Terminal Radar Approach Control
URET	User Request Evaluation Tool
WAK	Workload Assessment Keypad
WJHTC	William J. Hughes Technical Center
ZBW	Boston ARTCC
ZDC	Washington, DC, ARTCC
ZNY	New York ARTCC
ZOB	Cleveland ARTCC

Appendix A - Informed Consent Statement

P#_____ Date_______________
Airspace Type: En Route Terminal

I, _______________________________, understand that this study, entitled "Evaluation of Alternate Procedures for Departures in New York En Route and Terminal Airspace" is sponsored by the Federal Aviation Administration and is being directed by Dr. Todd R. Truitt.

Nature and Purpose:

I have been recruited to volunteer as a participant in this project. The purpose of the study is to determine the effects of alternative air traffic control procedures in a high-fidelity, controller-in-the-loop simulation. The results of the study will be used to establish the feasibility of implementing these alternative or similar air traffic control procedures in an operational environment.

Experimental Procedures:

En route Certified Professional Controllers (CPCs) will arrive at the simulation laboratory in groups of four and will participate over a three-day simulation session. Terminal CPCs will arrive at the simulation laboratory in groups of six and will participate over a one and a half day simulation session. Each participant will work complex traffic scenarios that involve handoffs with other participants. The first day of the simulation will consist of a project briefing, equipment familiarization, and four 50 min scenarios. The second day of the simulation will consist of four 50 min test scenarios. On the second day, en route CPCs will work all four scenarios and terminal CPCs will work two scenarios. On the third day, both en route and terminal CPC will work four 50 min scenarios. Participants will work from about 8:30 AM to about 5:00 PM every day with a lunch break and rest breaks after each traffic scenario.

Participants will control traffic under each of three different experimental procedures. After each scenario, participants will complete questionnaires to evaluate the impact of the alternative procedures on participant workload and acceptance. In addition, subject-matter experts will make over-the-shoulder observations during the simulation to further assess the procedures. Finally, an automated data collection system will record system operations and generate a set of standard ATC simulation measures, which include safety, capacity, efficiency, and communications measures. The simulation will be audio-video recorded in case researchers need to reexamine any important simulation events.

Discomfort and Risks:

I understand that I will not be exposed to any foreseeable risks or intrusive measurement techniques.

Confidentiality:

My participation is strictly confidential, and no individual names or identities will be recorded, associated with data, or released in any reports.

Benefits:

I understand that the only benefit to me is that I will be able to provide the researchers with valuable feedback and insight into the effects of alternative ATC procedures for departures from New York airspace. My data will help the FAA to establish the feasibility of these procedures within such an environment.

P#_____ Date_______________
Airspace Type: En Route Terminal

Participant Responsibilities:

I am aware that to participate in this study I must be a certified professional controller who is qualified at my facility and holds a current medical certificate. I will control traffic and answer any questions asked during the study to the best of my abilities. I will not discuss the content of the experiment with anyone until the study is completed.

Participant Assurances:

I understand that my participation in this study is completely voluntary and I have the freedom to withdraw at any time without penalty. I also understand that the researchers in this study may terminate my participation if they feel this to be in my best interest. I understand that if new findings develop during the course of this research that may relate to my decision to continue participation, I will be informed.

I have not given up any of my legal rights or released any individual or institution from liability for negligence.

Dr. Truitt has adequately answered all the questions I have asked about this study, my participation, and the procedures involved. I understand that Dr. Truitt or another member of the research team will be available to answer any questions concerning procedures throughout this study.

If I have questions about this study or need to report any adverse effects from the research procedures, I will contact Dr. Truitt at (609) 485-4351.

Compensation and Injury:

I agree to immediately report any injury or suspected adverse effect to Dr. Todd R. Truitt at (609) 485-4351. Local clinics and hospitals will provide any treatment, if necessary. I agree to provide, if requested, copies of all insurance and medical records arising from any such care for injuries/medical problems.

Signature Lines:

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

Research Participant:____ Date:__________
Investigator:____ Date:__________
Witness:____ Date:__________

Appendix B – Biographic Questionnaire

P#_____ **Date**_______________
Airspace Type: En Route Terminal

Instructions:

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

Biographic Information and Experience

1. What is your **gender**?	○ Male	○ Female

2. Will you be **wearing corrective lenses** during this experiment?	○ Yes	○ No

3. What is your **age**?	_____ years
4. How long have you worked as a **CPC (include both FAA and military experience)**?	_____ years _____ months
5. How long have you worked as a **CPC for the FAA**?	_____ years _____ months
6. How long have you been a **Certified Professional Controller** (or Full Performance Level Controller)?	_____ years _____ months
7. How long have you **actively controlled traffic** in the en route environment?	_____ years _____ months
8. How long have you **actively controlled traffic** in the terminal environment?	_____ years _____ months
9. How many of the **past 12 months** have you actively controlled traffic?	_____ months

P#_____ **Date**_______________
Airspace Type: En Route Terminal

10. Rate your current **skill as a CPC**.	Not Skilled	①②③④⑤⑥⑦⑧⑨⑩	Extremely Skilled

11. Rate your current **level of stress**.	Not Stressed	①②③④⑤⑥⑦⑧⑨⑩	Extremely Stressed

12. Rate your **level of motivation** to participate in this study.	Not Motivated	①②③④⑤⑥⑦⑧⑨⑩	Extremely Motivated

Appendix C – NASA-Task Load Index

P#_____ **Date**_______________ **Condition:** Normal Collocated Terminalized **Scenario:**______
Airspace Type: En Route Terminal **Sector:** 39 39H 55 55H WEST WESTH EWR EWRH LGA LGAH

Instructions:

Please answer the following questions based upon your experience in the scenario just completed. Your identity will remain anonymous.

Mental Demand – how much mental and perceptual activity was required (thinking, deciding, calculating, remembering, looking, searching, etc.)? Was your tasks easy or demanding, simple or complex, exacting or forgiving?

Rate your **mental demand** during this scenario.	Extremely Low	①②③④⑤⑥⑦⑧⑨⑩	Extremely High

Physical Demand – how much physical activity was required (e.g., data entry, strip marking, talking, pointing, etc.)? Was your task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Rate your **physical demand** during this scenario.	Extremely Low	①②③④⑤⑥⑦⑧⑨⑩	Extremely High

Temporal Demand – how much time pressure did you feel due to the rate or pace at which your tasks occurred? Was the pace slow and leisurely or rapid and frantic?

Rate your **temporal demand** during this scenario.	Extremely Low	①②③④⑤⑥⑦⑧⑨⑩	Extremely High

Performance – how successful do you think you were in accomplishing the goals of your tasks? How satisfied were you with your performance in accomplishing these goals?

Rate your **performance** during this scenario.	Extremely Low	①②③④⑤⑥⑦⑧⑨⑩	Extremely High

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

Rate your **effort** during this scenario.	Extremely Low	①②③④⑤⑥⑦⑧⑨⑩	Extremely High

Frustration – how insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel in performing your tasks?

Rate your **frustration** during this scenario.	Extremely Low	①②③④⑤⑥⑦⑧⑨⑩	Extremely High

Do you have any comments or clarifications about these NASA-TLX ratings?

Appendix D - Post-Scenario Questionnaire

P#_____ **Date**________________ **Condition:** Normal Collocated Terminalized **Scenario:**______
Airspace Type: En Route Terminal **Sector:** 39 39H 55 55H WEST WESTH EWR EWRH LGA LGAH

Instructions:
Please answer the following questions based upon your experience in the scenario just completed. Your identity will remain anonymous.

1. Rate your **overall level of ATC performance** during this scenario.	Extremely Poor	①②③④⑤⑥⑦⑧⑨⑩	Extremely Good
2. Rate your **ability to move aircraft through the sector** during this scenario.	Extremely Poor	①②③④⑤⑥⑦⑧⑨⑩	Extremely Good
3. Rate your **overall level of situation awareness** during this scenario.	Extremely Poor	①②③④⑤⑥⑦⑧⑨⑩	Extremely Good
4. Rate your **situation awareness for current aircraft locations** during this scenario.	Extremely Poor	①②③④⑤⑥⑦⑧⑨⑩	Extremely Good
5. Rate your **situation awareness for projected aircraft locations** during this scenario.	Extremely Poor	①②③④⑤⑥⑦⑧⑨⑩	Extremely Good
6. Rate your **situation awareness for potential aircraft loss-of-separation** during this scenario.	Extremely Poor	①②③④⑤⑥⑦⑧⑨⑩	Extremely Good
7. Rate your **workload due to air-to-ground communications** during this scenario.	Extremely Low	①②③④⑤⑥⑦⑧⑨⑩	Extremely High
8. Rate your **workload due to ground-to-ground communications** during this scenario.	Extremely Low	①②③④⑤⑥⑦⑧⑨⑩	Extremely High
9. Rate the **performance of the simulation pilots** in terms of their responding to your control instructions and providing readbacks.	Extremely Poor	①②③④⑤⑥⑦⑧⑨⑩	Extremely Good
10. Rate the **difficulty** of this scenario.	Extremely Easy	①②③④⑤⑥⑦⑧⑨⑩	Extremely Difficult

P#_____ **Date**_______________ **Condition:** Normal Collocated Terminalized **Scenario:**______
Airspace Type: En Route Terminal **Sector:** 39 39H 55 55H WEST WESTH EWR EWRH LGA LGAH

If you just completed the "**Terminalized**" condition, please answer the following question:

11. What effect, if any, did the reduced lateral separation standards have on your **ability to control traffic?**	Negative Effect	①②③④⑤⑥⑦⑧⑨ None	Positive Effect

Explain how the reduced lateral separation standards affected your **ability to control traffic**, if at all.

____

____

____

____

____

12. Were you **able to adapt** to the reduced lateral separation standards?	Not At All	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal

Explain **how you adapted** to the reduced lateral separation standards, if at all.

____

____

____

____

____

P#_____ **Date**_______________ **Condition:** Normal Collocated Terminalized **Scenario:**______
Airspace Type: En Route Terminal **Sector:** 39 39H 55 55H WEST WESTH EWR EWRH LGA LGAH

13. Do you have any additional comments or clarifications about your experience in the simulation?

____

____

____

____

____

____

____

____

____

Appendix E - Post-Experiment Questionnaire

P#_____ **Date**________________
Airspace Type: En Route Terminal

Instructions:

Please answer the following questions based upon your overall experience in the simulation. Your answers will remain anonymous.

1. Did your **communication strategies** change during the **collocated** condition (*without* reduced lateral separation standard)?	Not At All	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ ⑩	A Great Deal

Explain how **collocation** alone affected your **communication strategies**, if at all.

____
____
____
____
____

2. What effect, if any, did **collocation** alone (*without* reduced lateral separation standard) have on your **control strategies**?	Negative Effect	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ | None	Positive Effect

Explain how **collocation** alone affected your **control strategies**, if at all.

____
____
____
____
____

P#_____ Date_______________
Airspace Type: En Route Terminal

3. Did your **communication strategies** change during the **terminalized** condition (collocated *with* reduced lateral separation standard)?	Not At All	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal

Explain how the **terminalized** condition affected your **communication strategies**, if at all.

____

____

____

____

____

4. What effect, if any, did the **terminalized** condition (collocated *with* reduced lateral separation standard) have on your **control strategies**?	Negative Effect	①②③④⑤⑥⑦⑧⑨ | None	Positive Effect

Explain how the **terminalized** condition affected your **control strategies**, if at all.

____

____

____

____

____

5. Rate the **realism of the overall simulation experience** compared to actual ATC operations.	Extremely Unrealistic	①②③④⑤⑥⑦⑧⑨⑩	Extremely Realistic

6. Rate the **realism of the simulation hardware** compared to actual equipment.	Extremely Unrealistic	①②③④⑤⑥⑦⑧⑨⑩	Extremely Realistic

7. Rate the **realism of the simulation software** compared to actual functionality.	Extremely Unrealistic	①②③④⑤⑥⑦⑧⑨⑩	Extremely Realistic

8. Rate the **realism of the simulation traffic scenarios** compared to actual NAS traffic.	Extremely Unrealistic	①②③④⑤⑥⑦⑧⑨⑩	Extremely Realistic

P#_____ Date_______________
Airspace Type: En Route Terminal

9. Rate the **realism of the simulation airspace** compared to actual NAS airspace.	Extremely Unrealistic	①②③④⑤⑥⑦⑧⑨⑩	Extremely Realistic

10. To what extent did the **WAK online workload rating technique interfere** with your ATC performance?	None At All	①②③④⑤⑥⑦⑧⑨⑩	A Great Deal

11. Do you have any comments or suggestions for improvement about our simulation capability?

____

____

____

____

____

____

____

____

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

____

____

____

____

____

Appendix F – Communication Score Sheet

P#s____________ **Date**_______________ **Condition:** Normal Collocated Terminalized
Scenario:______

Communication Type	ARD >>> 74	ARD <<< 74
Coordination (e.g., hold)		
Traffic Management		
Glance		
Speed		
Heading		
Altitude		
Request for control		
Point Out		
Non-verbal (pointing)		
Non-ATC		
Other		

Appendix G – En Route ATC Observer Rating Form

P#_____ Date_______________ Condition: Normal Collocated Terminalized **Scenario:______**
Sector: 39 39H 55 55H

INSTRUCTIONS

This form is designed to be used by supervisory air traffic control specialists to evaluate the effectiveness of controllers working in simulation environments. SATCSs will observe and rate the performance of controllers in several different performance dimensions using the scale below as a general purpose guide. Use the entire scale range as much as possible. You will see a wide range of controller performance. Take extensive notes on what you see. Do not depend on your memory. Write down your observations. Space is provided after each scale for comments. You may make preliminary ratings during the course of the scenario. However, wait until the scenario is finished before making your final ratings and remain flexible until the end when you have had an opportunity to see all the available behavior. At all times please focus on what you actually see and hear. This includes what the controller does and what you might reasonably infer from the actions of the pilots. Try to avoid inferring what you think may be happening. If you do not observe relevant behavior or the results of that behavior, then you may leave a specific rating blank. Also, please write down any comments that may help improve this evaluation form. Do not write your name on the form itself. You will not be identified by name. An observer code known only to yourself and the researchers conducting this study will be assigned to you. The observations you make do not need to be restricted to the performance areas covered in this form and may include other areas that you think are important.

ASSUMPTIONS

ATC is a complex activity that contains both observable and unobservable behavior. There are so many complex behaviors involved that no observational rating form can cover everything. A sample of the behaviors is the best that can be achieved, and a good form focuses on those behaviors that controllers themselves have identified as the most relevant in terms of their overall performance. Most controller performance is at or above the minimum standards regarding safety and efficiency. The goal of the rating system is to differentiate performance above this minimum. The lowest rating should be assigned for meeting minimum standards and also for anything below the minimum since this should be a rare event. It is important for the observer/rater to feel comfortable using the entire scale and to understand that all ratings should be based on behavior that is actually observed.

P#_____ Date________________ **Condition:** Normal Collocated Terminalized **Scenario:______**
Sector: 39 39H 55 55H

Rating Scale Descriptors

Remove this Page and keep it available while doing ratings

SCALE	QUALITY	SUPPLEMENTARY
1	**Least Effective**	Unconfident, Indecisive, Inefficient, Disorganized, Behind the power curve, Rough, Leaves some tasks incomplete, Makes mistakes
2	**Poor**	May issue conflicting instructions, Doesn't plan completely
3	**Fair**	Distracted between tasks
4	**Low Satisfactory**	Postpones routine actions
5	**High Satisfactory**	Knows the job fairly well
6	**Good**	Works steadily, Solves most problems
7	**Very Good**	Knows the job thoroughly, Plans well
8	**Most Effective**	Confident, Decisive, Efficient, Organized, Ahead of the power curve, Smooth, Completes all necessary tasks, Makes no mistakes

P#_____ Date_______________ **Condition:** Normal Collocated Terminalized **Scenario:______**
Sector: 39 39H 55 55H

I - MAINTAINING SAFE AND EFFICIENT TRAFFIC FLOW
II - MAINTAINING ATTENTION AND SITUATION AWARENESS
III – PRIORITIZING
IV – PROVIDING CONTROL INFORMATION
V – TECHNICAL KNOWLEDGE
VI – COMMUNICATING
Handoff Position – Efficiency of Communication and Coordination

P#_____ **Date**_______________ **Condition:** Normal Collocated Terminalized **Scenario:**______
Sector: 39 39H 55 55H

I - MAINTAINING SAFE AND EFFICIENT TRAFFIC FLOW								
1. Maintaining Separation and Resolving Potential Conflicts .... • using control instructions that maintain appropriate aircraft and airspace separation • detecting and resolving impending conflicts early • recognizing the need for speed restrictions and wake turbulence separation	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
2. Sequencing Aircraft Efficiently .... • using efficient and orderly spacing techniques for arrival, departure, and en route aircraft • maintaining safe arrival and departure intervals that minimize delays	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
3. Using Control Instructions Effectively/Efficiently .... • providing accurate navigational assistance to pilots • issuing economical clearances that result in need for few additional instructions to handle aircraft completely • ensuring clearances require minimum necessary flight path changes	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
4. Overall Safe and Efficient Traffic Flow Scale Rating	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
II - MAINTAINING ATTENTION AND SITUATION AWARENESS								
5. Maintaining Awareness of Aircraft Positions .... • 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	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
6. Giving and Taking Handoffs in a Timely Manner .... • ensuring that handoffs are initiated in a timely manner • ensuring that handoffs are accepted in a timely manner • ensuring that handoffs are made according to procedures	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
7. Ensuring Positive Control .... • tailoring control actions to situation • using effective procedures for handling heavy, emergency, and unusual traffic situations	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
8. Detecting Pilot Deviations from Control Instructions .... • ensuring that pilots follow assigned clearances correctly • correcting pilot deviations in a timely manner	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
9. Correcting Own Errors in a Timely Manner.................................. • acting quickly to correct errors • changing an issued clearance when necessary to expedite traffic flow	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
10. Overall Attention and Situation Awareness Scale Rating	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**

P#_____ Date_______________ Condition: Normal Collocated Terminalized **Scenario:______**
Sector: 39 39H 55 55H

III – PRIORITIZING								
11. Taking Actions in an Appropriate Order of Importance .... • resolving situations that need immediate attention before handling low priority tasks • issuing control instructions in a prioritized, structured, and timely manner	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
12. Preplanning Control Actions .... • scanning adjacent sectors to plan for future and conflicting traffic • studying pending flight strips in bay	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
13. Handling Control Tasks for Several Aircraft................................. • shifting control tasks between several aircraft when necessary • communicating in timely fashion while sharing time with other actions	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
14. Marking Flight Strips while Performing Other Tasks .... • marking flight strips accurately while talking or performing other tasks • keeping flight strips current	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
15. Overall Prioritizing Scale Rating..	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
IV – PROVIDING CONTROL INFORMATION								
16. Providing Essential Air Traffic Control Information .... • providing mandatory services and advisories to pilots in a timely manner • exchanging essential information	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
17. Providing Additional Air Traffic Control Information.................. • providing additional services when workload permits • exchanging additional information	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
18. Providing Coordination .... • providing effective and timely coordination • using proper point-out procedures	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
19. Overall Providing Control Information Scale Rating.....................	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**

P#_____ **Date**_______________ **Condition:** Normal Collocated Terminalized **Scenario:**______
Sector: 39 39H 55 55H

V – TECHNICAL KNOWLEDGE								
20. Showing Knowledge of LOAs and SOPs • controlling traffic as depicted in current LOAs and SOPs • performing handoff procedures correctly	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
21a. Showing Knowledge of Aircraft Capabilities and Limitations • using appropriate speed, vectoring, and/or altitude assignments to separate aircraft with varied flight capabilities • issuing clearances that are within aircraft performance parameters	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
21b. Showing Effective Use of Equipment • updating data blocks • using equipment capabilities	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
22. Overall Technical Knowledge Scale Rating	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
VI – COMMUNICATING								
23. Using Proper Phraseology • using words and phrases specified in the 7110.65 • using phraseology that is appropriate for the situation • using minimum necessary verbiage	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
24. Communicating Clearly and Efficiently • speaking at the proper volume and rate for pilots to understand • speaking fluently while scanning or performing other tasks • ensuring clearance delivery is complete, correct and timely • speaking with confident, authoritative tone of voice	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
25. Listening to Pilot Readbacks and Requests • correcting pilot readback errors • acknowledging pilot or other controller requests promptly • processing requests correctly in a timely manner	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
26. Overall Communicating Scale Rating	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**

P#_____ **Date**_______________ **Condition:** Normal Collocated Terminalized **Scenario:**______
Sector: 39 39H 55 55H

Task	Frequency of Occurrence: Never Occurred	Rarely Occurred	Occurred, but within Normal Limits of Operational Acceptability	Occurred More Than Normal	Occurred Unacceptably Often
1. Errors or omissions in required flight strip marking	①	②	③	④	⑤
2. Gave arriving aircraft descent too early	①	②	③	④	⑤
3. Gave departing aircraft climb too late	①	②	③	④	⑤
4. Issued clearances earlier or later than appropriate	①	②	③	④	⑤
5. Failed to comply with Letters of Agreement	①	②	③	④	⑤
6. Offered handoffs earlier than appropriate	①	②	③	④	⑤
7. Offered handoffs later than appropriate	①	②	③	④	⑤
8. Accepted handoffs later than appropriate	①	②	③	④	⑤

Task	Overall ATC Performance: Operations were Unsafe and Unacceptable	Operational Safety was not Compromised, but I had Safety Concerns	Operations were Typical of this Type of Sector with Acceptable Safety	The Margin of Safety was Higher Than Normal for this Type of Sector
9. Overall Operational Assessment of ATC Performance	①	②	③	④

If you marked ① or ② for your overall operational assessment of ATC performance, please explain your rating below. Thoroughly describe the incidents or factors that influenced your judgment.

____

____

____

P#_____ **Date**_______________ **Condition:** Normal Collocated Terminalized **Scenario:**______
Sector: 39 39H 55 55H

Appendix H – Terminal ATC Observer Rating Form

P#_____ Date_______________ Condition: Normal Collocated Terminalized **Scenario:______**
Sector: WEST WESTH EWR EWRH LGA LGAH

INSTRUCTIONS

This form is designed to be used by supervisory air traffic control specialists to evaluate the effectiveness of controllers working in simulation environments. SATCSs will observe and rate the performance of controllers in several different performance dimensions using the scale below as a general purpose guide. Use the entire scale range as much as possible. You will see a wide range of controller performance. Take extensive notes on what you see. Do not depend on your memory. Write down your observations. Space is provided after each scale for comments. You may make preliminary ratings during the course of the scenario. However, wait until the scenario is finished before making your final ratings and remain flexible until the end when you have had an opportunity to see all the available behavior. At all times please focus on what you actually see and hear. This includes what the controller does and what you might reasonably infer from the actions of the pilots. Try to avoid inferring what you think may be happening. If you do not observe relevant behavior or the results of that behavior, then you may leave a specific rating blank. Also, please write down any comments that may help improve this evaluation form. Do not write your name on the form itself. Your identity will remain anonymous, as your data will be identified by an observer code known only to yourself and the researchers conducting this study. The observations you make do not need to be restricted to the performance areas covered in this form and may include other areas that you think are important.

ASSUMPTIONS

ATC is a complex activity that contains both observable and unobservable behavior. There are so many complex behaviors involved that no observational rating form can cover everything. A sample of the behaviors is the best that can be achieved, and a good form focuses on those behaviors that controllers themselves have identified as the most relevant in terms of their overall performance. Most controller performance is at or above the minimum standards regarding safety and efficiency. The goal of the rating system is to differentiate performance above this minimum. The lowest rating should be assigned for meeting minimum standards and also for anything below the minimum since this should be a rare event. It is important for the observer/rater to feel comfortable using the entire scale and to understand that all ratings should be based on behavior that is actually observed.

P#_____ **Date**________________ **Condition:** Normal Collocated Terminalized **Scenario:**______
Sector: WEST WESTH EWR EWRH LGA LGAH

Rating Scale Descriptors

Remove this Page and keep it available while doing ratings

SCALE	QUALITY	SUPPLEMENTARY
1	**Least Effective**	Unconfident, Indecisive, Inefficient, Disorganized, Behind the power curve, Rough, Leaves some tasks incomplete, Makes mistakes
2	**Poor**	May issue conflicting instructions, Doesn't plan completely
3	**Fair**	Distracted between tasks
4	**Low Satisfactory**	Postpones routine actions
5	**High Satisfactory**	Knows the job fairly well
6	**Good**	Works steadily, Solves most problems
7	**Very Good**	Knows the job thoroughly, Plans well
8	**Most Effective**	Confident, Decisive, Efficient, Organized, Ahead of the power curve, Smooth, Completes all necessary tasks, Makes no mistakes

P#_____ **Date**_______________ **Condition:** Normal Collocated Terminalized **Scenario:**______
Sector: WEST WESTH EWR EWRH LGA LGAH

I - MAINTAINING SAFE AND EFFICIENT TRAFFIC FLOW
II - MAINTAINING ATTENTION AND SITUATION AWARENESS
III – PRIORITIZING
IV – PROVIDING CONTROL INFORMATION
V – TECHNICAL KNOWLEDGE
VI – COMMUNICATING
Handoff Position – Efficiency of Communication and Coordination

P#_____ Date_______________ Condition: Normal Collocated Terminalized **Scenario:______**
Sector: WEST WESTH EWR EWRH LGA LGAH

I - MAINTAINING SAFE AND EFFICIENT TRAFFIC FLOW

1. Maintaining Separation and Resolving Potential Conflicts............ **1 2 3 4 5 6 7 8**
   - using control instructions that maintain appropriate aircraft and airspace separation
   - detecting and resolving impending conflicts early
   - recognizing the need for speed restrictions and wake turbulence separation

   Comments:

2. Sequencing Aircraft Efficiently.. **1 2 3 4 5 6 7 8**
   - using efficient and orderly spacing techniques for arrival and departure aircraft
   - maintaining safe arrival and departure intervals that minimize delays

   Comments:

3. Using Control Instructions Effectively/Efficiently......................... **1 2 3 4 5 6 7 8**
   - providing accurate navigational assistance to pilots
   - issuing economical clearances that result in need for few additional instructions to handle aircraft completely
   - ensuring clearances use minimum necessary flight path changes

   Comments:

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

   Comments:

P#_____ Date_______________ Condition: Normal Collocated Terminalized **Scenario:______**
Sector: WEST WESTH EWR EWRH LGA LGAH

II - MAINTAINING ATTENTION AND SITUATION AWARENESS

5. 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:

6. Giving and Taking Handoffs in a Timely Manner **1 2 3 4 5 6 7 8**
   - ensuring that handoffs are initiated in a timely manner
   - ensuring that handoffs are accepted in a timely manner
   - ensuring that handoffs are made according to procedures

   Comments:

7. Ensuring Positive Control.. **1 2 3 4 5 6 7 8**
   - tailoring control actions to situation
   - using effective procedures for handling heavy, emergency, and unusual traffic situations

   Comments:

8. Detecting Pilot Deviations from Control Instructions..................... **1 2 3 4 5 6 7 8**
   - ensuring that pilots follow assigned clearances correctly
   - correcting pilot deviations in a timely manner

   Comments:

9. Correcting Own Errors in a Timely Manner **1 2 3 4 5 6 7 8**
   - acting quickly to correct errors
   - changing an issued clearance when necessary to expedite traffic flow

   Comments:

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

    Comments:

P#_____ Date_______________ Condition: Normal Collocated Terminalized **Scenario:______**
Sector: WEST WESTH EWR EWRH LGA LGAH

III – PRIORITIZING

11. Taking Actions in an Appropriate Order of Importance **1 2 3 4 5 6 7 8**
    - resolving situations that need immediate attention before handling low priority tasks
    - issuing control instructions in a prioritized, structured, and timely manner

    Comments:

12. Preplanning Control Actions **1 2 3 4 5 6 7 8**
    - scanning adjacent sectors to plan for future and conflicting traffic
    - studying pending flight strips in bay

    Comments:

13. Handling Control Tasks for Several Aircraft **1 2 3 4 5 6 7 8**
    - shifting control tasks between several aircraft when necessary
    - communicating in timely fashion while sharing time with other actions

    Comments:

14. Marking Flight Strips while Performing Other Tasks **1 2 3 4 5 6 7 8**
    - marking flight strips accurately while talking or performing other tasks
    - keeping flight strips current

    Comments:

15. Overall Prioritizing Scale Rating... **1 2 3 4 5 6 7 8**

    Comments:

P#_____ **Date**_______________ **Condition:** Normal Collocated Terminalized **Scenario:**______
Sector: WEST WESTH EWR EWRH LGA LGAH

IV – PROVIDING CONTROL INFORMATION

16. Providing Essential Air Traffic Control Information **1 2 3 4 5 6 7 8**
    - providing mandatory services and advisories to pilots in a timely manner
    - exchanging essential information

    Comments:

17. Providing Additional Air Traffic Control Information.................... **1 2 3 4 5 6 7 8**
    - providing additional services when workload is not a factor
    - exchanging additional information

    Comments:

18. Providing Coordination **1 2 3 4 5 6 7 8**
    - providing effective and timely coordination
    - using proper point-out procedures

    Comments:

19. Overall Providing Control Information Scale Rating......................**1 2 3 4 5 6 7 8**

    Comments:

V – TECHNICAL KNOWLEDGE

20. Showing Knowledge of LOAs and SOPs.. **1 2 3 4 5 6 7 8**
    - controlling traffic as depicted in current LOAs and SOPs
    - performing handoff procedures correctly

    Comments:

21. Showing Knowledge of Aircraft Capabilities and Limitations **1 2 3 4 5 6 7 8**
    - using appropriate speed, vectoring, and/or altitude assignments to separate aircraft with varied flight capabilities
    - issuing clearances that are within aircraft performance parameters

    Comments:

22. Overall Technical Knowledge Scale Rating.................................... **1 2 3 4 5 6 7 8**

    Comments:

VI – COMMUNICATING

23. Using Proper Phraseology **1 2 3 4 5 6 7 8**
    - using words and phrases specified in the 7110.65
    - using phraseology that is appropriate for the situation
    - using minimum necessary verbiage

    Comments:

24. 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
    - ensuring clearance delivery is complete, correct and timely
    - speaking with confident, authoritative tone of voice

    Comments:

P#_____ **Date**_______________ **Condition:** Normal Collocated Terminalized **Scenario:**______
Sector: WEST WESTH EWR EWRH LGA LGAH

25. Listening to Pilot Readbacks and Requests **1 2 3 4 5 6 7 8**
    - correcting pilot readback errors
    - acknowledging pilot or other controller requests promptly
    - processing requests correctly in a timely manner

    Comments:

26. Overall Communicating Scale Rating **1 2 3 4 5 6 7 8**

    Comments:

27. Handoff Position – Communication and Coordination **1 2 3 4 5 6 7 8**

Appendix I

Comments on Data Analysis and the Repeated Measures Experimental Design

Experimenters often use a repeated measures design to control, and thereby reduce, the error variability in the data due to differences between participants. Too much error variability may prevent the researcher from detecting significant effects of experimental conditions (treatments). However, we must consider some special statistical assumptions when analyzing data from a repeated measures design. In a repeated measures design, the experimenter has set up the conditions such that participants in certain parts of the experiment are more alike than participants in other parts of the experiment. For example, participants who have expertise in one technical specialty are more similar to one another than to participants in a different technical specialty. Therefore, given repeated measurements, there is a correlation between the scores of participants in the same group (i.e., similar technical specialty and area-specific knowledge). The correlation of scores among participants also results in dependencies among experimental conditions.

Researchers initially justified the use of the *F* test in repeated measures designs by assuming that the condition of compound symmetry exists across conditions or participants. However, for the condition of compound symmetry to be met, each treatment must have the same true variance over all conditions (pooled within-group) and the covariances (across participants) for each pair of treatments must be a constant. While the assumption of compound symmetry is sufficient to justify the use of the *F* test[2] in a repeated measures design, it is not a necessary condition. In fact, the compound symmetry assumption is very strict and not likely to hold true, especially in experiments using a repeated measures design. The compound asymmetry assumption does not have to be met though in order to justify use of the *F* test. Huynh and Feldt (1970) and Rouanet and Lepine (1970), among others, have shown that the circularity assumption (or sphericity assumption), which is both mathematically necessary and sufficient, can be made to support the use of the *F* test in repeated measures designs. The circularity assumption simply states that the components of the within-subjects model are orthogonal (independent) components. For more information on the assumptions associated with repeated measures designs, refer to Kirk (1982) and Hays (1988).

One way to ensure that the statistical assumptions associated with a repeated measures design are satisfied is to analyze the data using the multivariate analysis of variance (MANOVA) method. In the MANOVA method, the different scores from each participant are handled as if they are actually scores from different variables. This method alleviates the necessity of the assumptions associated with the analysis of variance (ANOVA) *F* test. Significant MANOVA effects are then tested further by ANOVA *F* tests and particular post hoc comparisons. However, the MANOVA approach may not be feasible for small N designs where degrees of freedom are insufficient.

Another way to analyze data from a repeated measures design while accounting for the circularity assumption is to implement a three-step testing method as suggested by Kirk (1982) and Hays (1988). In this method, the data is first analyzed by an ANOVA. If the result is not significant then the analysis stops and one must conclude that there is no effect of the independent variables in question. If the ANOVA is significant, then the Geisser-Greenhouse (G-G) *F* test, or conservative *F* test, is conducted (Geisser & Greenhouse, 1958). Essentially, the

[2] The *F* test is justified (i.e., valid) when the reported *F* values adhere to the *F* distribution.

G-G *F* test adjusts the degrees of freedom used to calculate the *F* statistic to make the test more conservative (i.e., less likely to find a significant difference by chance where none exists). The G-G *F* test ensures that the researcher is not capitalizing on chance or on violations of the circularity assumption. If the G-G *F* test is significant, then the result is highly significant. If the G-G *F* test is not significant, then the circularity assumption may have been violated and the Box adjustment (Huynh-Feldt (H-F) *F* test or adjusted *F* test) is calculated (Huynh & Feldt, 1970). If the H-F *F* test is computed, then that result is the final determinant regarding whether a significant effect is present or not. We used this later method for the present experiment. We conducted multiple comparisons of means using Tukey's Honestly Significant Difference (HSD) test. If a significant interaction of main effects was found, then the Tukey's HSD test was computed to explain the interaction for all relevant analyses.

We selected this three-step approach to minimize the probability of a Type II error (i.e., false acceptance of the null hypothesis or finding no effect where one actually exists) while sacrificing an increase in the probability of Type I error (i.e., false rejection of the null hypothesis or finding an effect where none actually exists). Such an approach will increase the likelihood the statistical analyses will detect any potentially detrimental effects caused by the Collocated or Terminalized conditions. We balanced this arguably liberal approach to data analysis by using the Tukey's HSD to conduct post hoc tests rather than calculating simple main effects.

Appendix J – Instruction Set

First Experimental Scenario

Equipment Familiarization (Practice)

During this brief practice scenario, please take the opportunity to familiarize yourself with your radar or handoff position. Familiarize yourself with the landlines and the Workload Assessment Keypads, or WAKs as we call them. This practice scenario is for your benefit and you should use this time to prepare for the scenarios that will follow. I will now read the WAK instructions to you.

Normal Condition Instructions

During this scenario, we would like you to control traffic as you normally would in the field. As in every scenario, you will be making workload ratings using the WAK. I will now read the WAK instructions to you.

Collocated Condition Instructions

During this scenario, we will simulate the collocation of terminal and en route facilities. Because there is no physical barrier between the terminal and en route sectors, face-to-face communication is possible and you may use it at your discretion. As in every scenario, you will be making workload ratings using the WAK. I will now read the WAK instructions to you.

Terminalized Condition Instructions

During this scenario, we will simulate the collocation of terminal and en route facilities. Because there is no physical barrier between the terminal and en route sectors, face-to-face communication is possible and you may use it at your discretion. In addition, Sector 74, Broadway, will use an alternative separation standard. Sector 74's lateral separation standard will be reduced from 5 nautical miles to 3 nautical miles. The halos and conflict alert algorithm will be adjusted accordingly As in every scenario, you will be making workload ratings using the WAK. I will now read the WAK instructions to you.

WAK Instructions

One purpose of this research is to obtain an accurate evaluation of controller workload. By workload, we mean all the physical and mental effort that you must exert to do your job. This includes maintaining the "picture," planning, coordinating, decision making, communicating, and whatever else is required to maintain a safe and expeditious traffic flow. Every five minutes the WAK device, located at your position, will emit a brief tone and the ten buttons will illuminate. The buttons will remain lit for only a limited amount of time. Please tell us how hard you are working at that moment by pushing one of the buttons numbered from 1 to 10.

I will review what these buttons mean in terms of your workload. At the low end of the scale (1 or 2), your workload is low - you can accomplish everything easily. As the numbers increase, your workload is getting higher. The numbers 3, 4, and 5 represent increasing levels of moderate workload where the chance of error is still low but steadily increasing. The numbers 6, 7, and 8 reflect relatively high workload where there is some chance of making errors. At the high end of the scale are the numbers 9 and 10, which represent a very high workload, where it is likely that you will have to leave some tasks unfinished.

All controllers, no matter how proficient and experienced, will be exposed at one time or another to all levels of workload. It does not detract from a controller's professionalism when he indicates that he is working very hard or that he is hardly working. Feel free to use the entire scale and tell us honestly how hard you are working. Do not sacrifice the safe and expeditious flow of traffic in order to respond to the WAK device. Remember, your workload rating should *not* reflect how much you are working during the course of the scenario. Instead, your rating should reflect how much workload you are experiencing during the instant when you are prompted to make the rating.

Does anyone have any questions? *(After answering questions, if any, instruct participants to do comm check with pilots and adjacent sectors and centers.)*

Instructions for Subsequent Scenarios

Equipment Familiarization (Practice)

During this brief practice scenario, please take the opportunity to familiarize yourself with your radar or handoff position. Familiarize yourself with the landlines and the Workload Assessment Keypads, or WAKs as we call them. This practice scenario is for your benefit and you should use this time to prepare for the scenarios that will follow. I will now read the WAK instructions to you.

Normal Condition Instructions

During this scenario, we would like you to control traffic as you normally would in the field. As in every scenario, you will be making workload ratings using the WAK. I will now read the WAK instructions to you.

Collocated Condition Instructions

During this scenario, we will simulate the collocation of terminal and en route facilities. Because there is no physical barrier between the terminal and en route sectors, face-to-face communication is possible and you may use it at your discretion. As in every scenario, you will be making workload ratings using the WAK. I will now read the WAK instructions to you.

Terminalized Condition Instructions

During this scenario, we will simulate the collocation of terminal and en route facilities. Because there is no physical barrier between the terminal and en route sectors, face-to-face communication is possible and you may use it at your discretion. In addition, Sector 74, Broadway, will use an alternative separation standard. Sector 74's lateral separation standard will be reduced from 5 nautical miles to 3 nautical miles. The halos and conflict alert algorithm will be adjusted accordingly As in every scenario, you will be making workload ratings using the WAK. I will now read the WAK instructions to you.

WAK Instructions

I will review what the WAK buttons mean in terms of your workload. At the low end of the scale (1 or 2), your workload is low - you can accomplish everything easily. As the numbers increase, your workload is getting higher. The numbers 3, 4, and 5 represent increasing levels of moderate workload where the chance of error is still low but steadily increasing. The numbers 6, 7, and 8 reflect relatively high workload where there is some chance of making errors. At the high end of the scale are the numbers 9 and 10, which represent a very high workload, where it is likely that you will have to leave some tasks unfinished.

All controllers, no matter how proficient and experienced, will be exposed at one time or another to all levels of workload. It does not detract from a controller's professionalism when he indicates that he is working very hard or that he is hardly working. Feel free to use the entire scale and tell us honestly how hard you are working. Do not sacrifice the safe and expeditious flow of traffic in order to respond to the WAK device. Remember, your workload rating should *not* reflect how much you are working during the course of the scenario. Instead, your rating should reflect how much workload you are experiencing during the instant when you are prompted to make the rating.

Does anyone have any questions? *(After answering questions, if any, instruct participants to do comm check with pilots and adjacent sectors and centers.)*

Appendix K – External Review

First NASA Review of "Effects of Collocation and Reduced Lateral Separation Standards in the New York Integrated Control Complex"

<u>Executive Summary of Review:</u>

This report describes a series of experiments ("the Study") to examine the potential benefits of collocating the New York Center (ZNY) and New York TRACON (N90) into one facility (the New York Integrated Control Complex, or NYICC).

While the overall Study itself was conducted with a reasonably appropriate application of methods and analysis, it is my overall summary that the Study did not investigate the "right" questions that would support the conclusions that are drawn. The authors conclude that there are significant benefits in an integrated Center/TRACON environment that can be attributed to improved communications between sectors and reduced lateral separation standards. The Study focused on some basic assumptions in examining these variables, which are not conclusive; further, I do not believe these assumptions are basic to improving operations anticipated by the NYICC concept. In addition, much of the analysis description does not include standard reporting of effect sizes (for example, η^2 or ω^2) with which to evaluate the results. It is up to the reader to interpret the relative importance of a reported effect, given the statistical data and the investigator's conclusions. Without effect size values, there are descriptions of trends, and the potential significance of trends that may be operationally significant, but whose statistical significance is unknown (these could be derived with a little work). While there is an excellent discussion of the problem in ATC research where findings are operationally but not statistically significant, the authors nonetheless draw conclusions based on findings in their data that may have occurred by chance, without warning readers of this fact. It is understandable that the authors wish to point out potential benefits that may not be borne about by statistical analysis, but this distinction can be made without what appears to be an attempt to bolster results supportive of the NYICC concept.

Response: We will address some of the issues in the Executive Summary separately in the more detailed comments below. However, we would like to respond directly to some of them now. First, it is important to not overlook the positive amongst the negative. The statement that overall the study used appropriate methods and analyses is crucial, in that there is nothing inherently and fatally flawed that will not allow us to make any use of the data. In addition, the reviewer found our discussion about distinguishing between operational and statistical significance to be excellent, which lays the foundation for the treatment of the operational data as opposed to the subjective data provided on questionnaires.

It is unclear exactly what the "right" questions are or what assumptions we made that are invalid, but the more detailed discussion indicates the right questions are related to other aspects of the NYICC concept (which we clearly indicated were not examined and that other changes could mitigate the obtained results) and that the assumptions are related to the impact of face-to-face (FTF) communications. We will discuss them more below, but simply want to state we are not 100% certain what the reviewer means by these summary comments.

In terms of not warning the reader about the potential for effects occurring by chance, we did state that marginal findings (p. 16) do not provide the same strength of support and assumed that data that was not analyzed statistically might be caused by chance. We will be glad to discuss that possibility more clearly but propose to also discuss evaluating the overall pattern of results as an indication that the effects are systematic rather than chance. Of course, the entire scientific method is probabilistic and never results in absolute proof; it's just that statistical treatments quantify the risk, on average. That is, a significance test at $p < .05$ simply means that a given result would have occurred solely by chance less than 5 times in 100 tests, but you never know whether any particular test was one of the five.

Finally, the reviewer's closing statement about "what appears to be an attempt to bolster results supportive of the NYICC concept." is unwarranted. We simply reported the findings and pointed out that they are supportive of the collocation and reduced separation aspects of the NYICC. If we had found reduced throughput or other negative outcomes in the two experimental conditions, we would have reported them and said they were not supportive. In fact, we did point out in our conclusion section that on a few variables, the Collocated condition was actually worse than the Normal condition. However, the overall pattern of results was in a positive direction for the Collocated condition and consistently so for the Terminalized condition. This overall pattern is a key to the conclusion; if the observed differences were due to chance, the most likely outcome is that roughly half of the variables would be in a positive direction and the other half would have been in a negative direction instead of almost always or always in a positive direction, respectively. It is theoretically possible that all the outcomes could be in one direction or the other, but extremely unlikely. Not once in the General Discussion and Conclusions did we say there was statistically significant support for either aspect, and we clearly qualified the supportive conclusion because of numerous causes, such as the impact of airspace design changes, ability of other elements of the NAS to handle the increased throughput, layout of the facility, etc.

Review:
Study Rationale

The NYICC Concept of Operations, as cited in the ACE Plan (2002) suggests that NYICC benefits include:

1. Reduce fragmentation of arrival and departure corridors across multiple centers, which currently limits flexibility;
2. Provide additional airspace under terminal control to support a more even balance of arrivals among arrival fixes and holding patterns within the TRACON; and
3. In general, improve capacity through reduced delays, reduced restrictions, and enhanced operations during severe weather.

Based on the Congressional testimony that was provided as part of the review materials, the NYICC is intended to: "allow air traffic controllers to better integrate traffic into the en route environment and more efficiently sequence traffic to a region, rather than an airport, allowing for more efficient airspace designs."

The Study concentrates on emphasizing the human factors impact of two major research questions:

i) Will a combined facility facilitate face to face communication between adjacent sectors (Center and TRACON) that will, in turn, contribute to improved operations, and,

ii) Will a combined facility that expands terminal-area separation standards to portions of the en-route environment result in improved operations?

There is a significant discrepancy between the assumed benefits from the NYICC Operational Concept and what the Study set out to determine.

Response: There are three areas of NYICC Operational Impacts cited in the Concept of Operations document (May 8, 2002): airspace environment, operational environment, and traffic handling. This study only addressed the operational environment, which includes collocating N90 and ZNY and reducing separation standards. The effects of collocation (see p. 12 of the Concept of Operations) were hypothesized to increase capacity and reduce delays as a result of physical proximity allowing improved situation awareness, minimizing procedural coordination, and making control decisions to accommodate the needs of both own sector and other sectors. The expansion of terminal airspace was hypothesized to enable controllers to safely increase throughput (see p. 13 of the Concept of Operations). These are the types of variables we measured while systematically manipulating the operational environment between separate facilities, a collocated facility, and collocated with terminal separation. We were not tasked to examine the other aspects of the NYICC concept, nor could we have in the time frame allowed to *conduct these simulations. For example, the redesign of the airspace was not complete, so there was no basis for changing the airspace. It would also confound the study of the operational environment because the controllers would be unfamiliar with it. It is normal in science to break down large, complex issues into smaller components and examine them under relatively controlled conditions, while recognizing that additional factors can influence the results. As a practical matter, it is virtually impossible to include every variable in a single study and be able to interpret the results. We believe the simulations we conducted directly addressed the issues we were tasked to evaluate, and are not significantly discrepant from the assumed benefits of the operational environment aspects of the NYICC Operational Concept.*

The basis for the NYICC concept is the well-studied and documented fact that the N90 and ZNY airspace environments are extremely complex. Communication difficulty (which has not been demonstrated by any literature review to be a true bottleneck in the efficiency of N90/ZNY operations) is but one artifact of the complexity of the operations in this airspace which involves the sequencing and spacing of arrival and departure traffic to and from at least four major airports (EWR, LGA, JFK, TEB), and deals with small sector sizes, significant crossing traffic, and flows to and from adjacent facilities such as Philadelphia, Boston, and the Washington, D.C. Area. The structure of the airspace in the Northeast was not designed with current levels of traffic in mind. Improvement from a combined facility assumes that operations can be managed from a single viewpoint, at the same time taking a more regional view, and enabling better control of the system and giving flexibility to the overall plan essentially by allowing one entity to do the planning. It was suggested, in an e-mail that solicited review of the Study, that there is significant congestion in N90/ZNY that somehow arises specifically from the burden of having controllers communicate between facilities. This statement is not true. Controllers are required to communicate between sectors and between facilities all over the NAS. It is not that the burden of communication is somehow irrationally large between N90 and ZNY. There is a burden that could arise, however, if the information provided between facilities is inconsistent or requires significant effort to resolve, or if different parties make decisions or responses at odds with one another using the same set of information and do not communicate their intent.

Response: It is absolutely true that the ZNY airspace is extremely complex and not designed to handle today's (much less tomorrow's) traffic. However, we were not tasked to evaluate the airspace, just the proposed collocation and separation standards aspects of the concept. The authors do not know if the communication congestion in the current environment represents a major burden or is irrationally large. In these two simulations, we evaluated the communications between the different conditions to test the hypotheses in the Concept of Operations (see pp. 7-8) that physical collocation would allow increased capacity resulting from greater situation awareness of interacting sectors and reducing 'proceduralized' coordination, which is perceived as more time consuming than informal coordination. By manipulating the operational environment and measuring both communications data and throughput data, we were testing whether these hypotheses were valid. Compared to the Normal (separate facilities) condition, both the Collocated and Terminalized conditions showed reduced landline communications, a greater number of FTF interactions involved in sharing relevant information, *and various measures of increased throughput (e.g., more arriving aircraft, more departing aircraft, less time in holding, and more aircraft handled).*

The Study's emphasis on terminalization examines only one potential by-product of collocating the two facilities: reduced lateral separation in the traditionally "en-route" environment. Realistically, collocating the operations of the two facilities will in fact need a careful examination of what airspace restructuring would have to occur, and how best to combine adjacent sector responsibilities. The Study should acknowledge that merely reducing separation standards without adjusting sectors in any way is a simplification of what could change in NYICC.

Response: We are not sure why the reviewer considers reduced separation in the en route environment to be a byproduct of collocation. We thoroughly agree and pointed out clearly in our General Discussion and Conclusion Section, that how a combined facility is designed must be carefully considered and that changes in the airspace may amplify or mitigate the effects of collocation or terminalization alone. The results we obtained relative to reducing separation were limited to no other adjustment in the sectors. Our qualification that airspace changes could affect the results is already acknowledged.

In addition, the terminalization condition is really confounded by including the collocation condition within it. It would have been a cleaner analysis and interpretation had the reduced separation standards been investigated as a condition on its own, and without the additional collocation condition included. Therefore it is difficult to interpret whether benefits from terminalization are really chiefly due to separation standards reduction or due to collocation, or a combination of the two. The Study can **not** suggest a clear effect due to reduced separation standards.

Response: We agree it would have been better to have had an additional condition in which en route airspace used terminal separation, but the facilities were not collocated. In fact, we proposed that experimental design in our second meeting with the Airspace Design Team, but the idea was rejected because it was not part of the Concept of Operations and because of the additional time and effort needed to design and run the expanded simulations. We do not agree that you cannot draw inferences about whether using terminal separation standards in a collocated facility produces different results than collocated only or in the current condition. We can and did make legitimate comparisons between these conditions. What you cannot determine from the experimental design is whether collocation is necessary to achieve the same (or even better) benefits from using terminal separation. If there were an alternative concept to terminalize en route airspace without collocating facilities, then that concept would require additional research.

The emphasis on communication is likewise another by-product of the potential benefits in a combined facility. There is an overall assumption that face-to-face communications are somehow inherently better than land line communications. One ATC-related reference (Rognin & Blanquart, 2000) is given as evidence that face-to-face communication would be beneficial. The quote that is included in the Study report actually comes from a study of D-side and R-side controllers. The benefits of collocating the D-side and R-side controllers are entirely different from collocating two radar controllers from the adjacent sectors of two different facilities. In the case of the D- and R-side controllers, these are two controllers that are collaborating on the same task, that of ensuring separation between aircraft in a single sector; they share the same rules and responsibilities in that regard. The radar controllers in adjacent sectors have different rules/procedures specific to their sectors and interact differently than two controllers sharing responsibilities in the same sector.

Response: In the Concept of Operations, procedural coordination via landline is cited as taking more time and limiting the number of aircraft a controller can handle. By enabling face-to-face interaction, collocation was hypothesized to reduce the negative aspects of procedural coordination. Our simulations tested these hypotheses by comparing controller behavior and system throughput under simulated normal (separate facilities) and collocated conditions. In the Rognin & Blanquart (2000) study, they discussed both between sector (both within a facility and between facilities) and within sector coordination between R and D controllers, but their observations focused on the en route R and D team. Some of the observations they made about being collocated were as follows. First, "co-located agents have the opportunity to observe each other, distributing and acquiring explicitly as well as implicitly information, through verbal messages, visual observation of other agents and of informational supports such as the radar screen, the strip progress board, the radio or the notepad. Thus, the working position provides some cues and informs about the current and planned actions and usually enables co-operators to infer their colleagues' current intentions and strategy." Second, they elaborated on the following benefits of awareness gained by physical proximity: awareness about who is talking, about the availability of the other, about current actions, and about the current situation. While working with interacting sectors is somewhat different than an R and D working a single sector, it seems that the advantages of collocation are similar for both. Just as the R and D have to coordinate their actions, so too do the separate sectors have to coordinate to handoff traffic, adjust flows, or request deviations from normal procedures.

There are potentially many drawbacks to the face-to-face communication argument. Face-to-face communications may draw controllers' visual scan away from the radar scope, a potential safety issue. There is similarly no mention of the fact that face-to-face communications are non-standard, easily misinterpreted, and not recorded, thereby reducing accountability.

Response: It is difficult to understand why the reviewer believes these drawbacks would occur when collocating terminal and en route sectors. In each separate facility today, the controllers that are collocated and interacting look at each others' radar displays, flight strips, etc. and coordinate face to face without it being recorded. That is, it's being done today in every facility without any known safety risks. Why would collocation of sectors from ZNY and N90 be any different than interacting sectors within each facility? Furthermore, we did not observe any of the negative effects hypothesized by the reviewer during the simulations.

Other factors contribute to effective communication that are unrelated to the location of the two communicating parties: the commonality of goals, the quality of data or information that is provided to the two parties, and the defined individual roles and responsibilities. Further, proper training and procedures may obviate any benefits due to collocation. These are complicating and mitigating factors in evaluating the impact of collocating adjacent sectors that should be acknowledged and which merit better investigation (what, for example, are the current complaints of the adjacent [ZNY and N90] sectors in terms of what communication problems they currently experience?). If anecdotal data that describes current-day communication problems were collected and described, that would have lent much more credence to the collocation hypothesis.

Response: We agree that the other factors cited that can affect communication are important; in fact, Rognin and Blanquart (2000) cite them explicitly. It is not clear how they are pertinent to the issue of collocation. Even the reviewer states they are unrelated to the location of the communicators, so we are not sure if this statement is a criticism of the report. We would not expect the commonality of goals or definitions of roles and responsibilities to change whether in one facility or two. It is hard to imagine that the quality of information could be worse if collocated than separated. It is also difficult to see how "proper training and procedures" could obviate any benefits of collocation. There are procedures in place for interfacility coordination, and presumably the controllers are trained to use them. According to the Concept of Operations, these procedures are time consuming and more difficult than face-to-face coordination. Although authors are not listed on the Concept document, we believe that controllers from ZNY and N90 contributed to it and thereby present the anecdotal evidence the reviewer seeks. In our simulations, we included a Normal condition, representing sectors from both ZNY and N90, so that we could compare communications (both objectively from the numbers of calls and types of FTF exchanges and subjectively in terms of participant questionnaires) with the Collocated and Terminalized conditions under otherwise similar traffic scenarios. If we did not have the Normal condition, then it would be more important to assess current-day communications in the facilities, but the underlying scenarios would differ (our scenarios were 130% traffic loads representing the future environment).

There are many additional questions to be answered from an NYICC concept. The authors do acknowledge that the application of the 3 nmi separation rule does not suggest that the Tower operations can handle the resulting increased traffic flow. In addition, there should be some discussion about the extrapolation to real operations, considering arrival flows into the other major N90 airports (LGA, JFK, TEB), and the impact on merging several flows of 30% more traffic with 3 nmi separations on final.

Response: We agree there are additional questions to be answered about the NYICC concept. We pointed out some of the more important ones in the General Discussion and Conclusions section of the report. Everything mentioned by the reviewer and more are noted there to caution the reader that while these two simulations provide data supportive of the collocation and reduced separation aspects of NYICC, it "does not mean that all issues associated with the implementation of these proposed changes have been resolved" (p. 103). The simulations we conducted were designed to meet the specific requirements of the sponsor within the constraints of cost, schedule, participant availability, laboratory capability, and experimental control. It is not practical to address all issues at once, but we were careful to qualify our conclusions by pointing out the potential for other issues to affect them either positively or negatively. This approach is standard scientific practice.

Methodology and Analysis

In general, the analyses (ANOVAs were used) appear to be appropriate and thorough. Specific concerns (that were alluded to above) include:

1. No reporting of effect sizes (for example, η^2 or ω^2) which allow the reader to better interpret the results.

Response: While the American Psychological Association Publication Manual (2001) recommends measures of effect size to increase the interpretability of statistical results, they are not required (see pp. 23 and 25). They are far from standard use, especially in applied research where the operational context assists in determining the importance of an effect more than the statistical measure. I skimmed the latest three issues of the journal, Human Factors, to see how widespread these measures were used. Of the 36 articles in the three issues, only 3 included magnitude of effect measures, and they all used eta squared (the first one recommended by the reviewer). One of the three was a study of mathematical modeling of decision making, so it is logical they would be more inclined to report additional mathematical information.

However, there are problems with the use of these measures (Tabachnick & Fidell, 1989; pp. 54-55). They note that the percentage of variance accounted for by an independent variable (IV) by eta squared depends on the number and significance of the other IVs in the design. That is, the *results would differ depending on whether there was a one-way or two-way test of the same measures. There is an alternate form of eta squared that addresses this problem, but the sum of the variances accounted for can sum to more than 100%. In addition, eta squared only estimates the proportion of the variance in the sample, not in the population to which the author wishes to draw inferences. Omega squared (the second measure suggested by the reviewer) was developed to estimate strength of association in the population. Unfortunately, its use is limited to between-subjects ANOVA with an equal number of participants in each cell. There is an alternate form of omega squared developed for repeated measures designs, in which separate measures are computed for each statistically significant main and interaction effect. For those measures that we analyzed statistically, most were position-by-condition tests, where position is a between-subjects IV and condition is a within-subjects IV. Even if these measures were computed, there is no absolute criteria for how large an effect must be for it to be meaningful. Finally, magnitude of effect cannot be measured for those variables for which statistical analyses were not appropriate. While we understand the rationale for estimates of effect sizes, we do not think they are appropriate or necessary for this study.*

American Psychological Association (2001). *Publication Manual of the American Psychological AssociationI (Fifth Edition).* Washington, DC: Author.

Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics (Second Edition).* New York: HarperCollinsPublishers, Inc.

2. From the long description in *Section 3.2, System Performance Measures,* the authors discuss the justification of reporting marginal, but non-significant results, and this is misleading. If the results are not statistically significant, but have potential importance, they may be reported with the caveat that those findings may have occurred by chance. There are many explanations as to why an effect may not be detected, and that is worth explaining (small sample size, a factor whose effect is subtle, and the confounding of the terminalization condition).

Response: There are several factors that can explain why an effect may not be detected, but certainly sample size is a major concern. There is an inherent tradeoff between the risks and consequences of two types of error in scientific studies. The first is the risk of concluding that an effect exists, when in fact the differences observed in an experiment were based on chance variation. This error is called alpha or Type I, and represents the concern of the reviewer. The other error, called beta or Type II, occurs when there is a real effect in the population, but the research fails to detect it. The smaller the sample size, the less likely the researcher is to find a significant difference in a test for an effect that is constant in the population, thus avoiding Type I errors but risking Type II errors. That is, the power of the statistical test to detect a real effect is inversely related to the sample size. The consequences of the two error types depend on the *nature of the research question. If the research fails to detect a harmful effect (e.g., an increase in workload that could lead to operational errors), the Type II error is the more consequential. If the statistical test indicates a positive effect (e.g., increased situation awareness that helped to avoid operational errors) that occurred by chance and would not apply in reality, then Type I is the more consequential error.*

In this pair of simulations, our sample size was limited because of costs, number of controllers certified on the sectors we simulated, and impact to the operations of ZNY and N90. Therefore, we took a two-pronged approach to significance testing on those variables that could be statistically evaluated. The first was to test against the most common, albeit arbitrary, criterion of $p < .05$ *(1 chance in 20 of a chance finding). In addition, we subjected all significant ANOVA test results to the more conservative Geisser-Greenhouse F test and Huynh-Feldt F test to ensure we did not violate any statistical assumptions and to increase our confidence we were not capitalizing on chance variation across variables. We never say an effect is statistically significant without passing all three of these tests. However, we did not want to miss detecting any real effects (Type II errors) because of the small sample size, so we set a secondary criterion of* $p < .10$ *(1 chance in 10 of a chance finding), which we called marginal results as a warning that these have a greater probability of being a chance occurrence, but the probability remains small. Then we looked across all the measures for patterns of results. We believe this approach minimizes the risks of both Type I and Type II errors under the constraints of limited sample size. In addition, although our absolute sample size was small (which affects statistical testing), we had a comparatively large percentage of the total relevant population participating in the simulations (which increases confidence in extrapolating to the larger population). The comparison in mind is academic research based on animals or college sophomores and then extrapolating to all humans or subsets thereof (male-female, young-middle-old, etc.).*

The discussion about marginal results was actually in Section 3.0, not 3.2.

3. Impact of small sample size on results. One of the Conclusions that was drawn was that the departure simulation effects were stronger than the arrival simulation effects, and this discrepancy was attributed to the greater complexity of the departure simulation (simulating more airports and flows). Clearly, this, if anything, should contribute to lower effects or benefits due to the introduction of greater random error. It is obvious that the larger sample size in the departure simulation contributed to more robust results, and the smaller sample size, and more simplistic flow in the arrival simulation was simply not enough to draw more substantive conclusions. The reader is unable to evaluate this, however, since effect sizes are not reported.

Response: The reviewer's comments appear to be based on the general approach of scientific research to take complex phenomena (e.g., learning) that occurs in complex environments *(school buildings, teachers, media, topics, other students, policies, home life, etc.) and breaking it down to small components where all but a few variables are controlled, the others systematically manipulated, and then a few measures of learning (number of trials to criteria, time to learn, etc.) are taken for analysis. We understand and use this methodology (e.g., we didn't include every sector and all the potential other changes to the airspace), although we do not, for practical reasons, break the research down into the smaller, more controlled experiments done in academic laboratories where they are researching theoretical issues. To do so would create an unacceptably artificial environment in which the results could not be generalized to actual operations. The corollary to this approach is that the more complex the experimental situation, the less likely a real effect will be detected in the test because of all the interacting and potentially confounding influences, or it will be more difficult to attribute an observed effect to a specific cause. In many cases, this is true.*

However, there is an alternative logic that we believe applies to the results we observed. When participants in a study are given a small set of options to perform under a current situation, there is a ceiling on how much improvement can be observed when the situation is changed. When more options are available and exercised judiciously, then more improvement is possible. The variables we were discussing regarding greater effects observed in the Departure simulation compared to the Arrival simulation were not subjected to statistical analyses because of the reasons cited in our report. They were the throughput measures that result not just from the actions of any one sector team, but through the efforts of all participating controllers and sectors. In the Arrival simulation, the primary flow of traffic was through Sector 75 to Sector 74 to ARD to EWR, a linear problem with limited options for increasing throughput, although there were some increases in the experimental conditions. In the Departure simulation, both EWR and LGA were simultaneously feeding aircraft to LIBW (along with other, nonsimulated airports) who then fed both Sectors 39 and 55, thus leading to more options for optimizing throughput. Although the variables were not always the same in the two simulations (e.g., airborne holding time in the Arrival simulation and ground delays in the Departure simulation), the throughput improvements observed in the Collocated and Terminalized conditions compared to the Normal condition were of a larger magnitude in the Departure simulation. Because there were no statistical tests of these variables, the issue of measures of effect size is not relevant.

Finally, the issue of their being a larger sample size (number of participants) in the Departure simulation ($n = 32$ versus $n = 18$ in the Arrival simulation) is not as straightforward as it appears. Sample size is normally associated with the number of participants on the assumption they will each contribute independent measures on a variable that are then tested. That is not true in this situation because, as noted above, the throughput measures were based on scenarios across positions, not within a position. On the en route sectors, we had 9 controllers in the Arrival simulation and only 8 in the Departure simulation. However, each controller participated in the three experimental conditions in three positions (two radar, one handoff) in the first simulation but in four positions (two radar, two handoff) in the second. In the terminal sectors, we had 9 controllers in the Arrival simulation who worked three scenarios in each of three positions (two radar, one handoff). We had 24 controllers in the Departure simulation *who worked three scenarios in only two positions (one radar, one handoff). These variations were based on which controllers were qualified to work which sectors.*

When counting the number of data points in each experimental condition (Normal, Collocated, and Terminalized) that provided data for the throughput measures, there were 9 runs each in the Arrival simulation and 8 runs each in the Departure simulation. Therefore, the number of data points was actually larger in the Arrival simulation, but only by one scenario run.

4. Lack of ability to really interpret the terminalization results because they contain the collocation and the reduced separation standard in the same condition.

Repeat Response: We agree it would have been better to have had an additional condition in which en route airspace used terminal separation, but the facilities were not collocated. In fact, we proposed that experimental design in our second meeting with the Airspace Design Team, but the idea was rejected as not being part of the Concept of Operations and because of the additional time and effort needed to design and run the expanded simulations. We do not agree that you cannot draw inferences about whether using terminal separation standards in a collocated facility produces different results than collocated only or in the current condition. We can and did make legitimate comparisons between the conditions. What you cannot determine from the experimental design is whether collocation is necessary to achieve the same (or even better) benefits from using terminal separation. If there were an alternative concept to terminalize en route airspace without collocating facilities, then that situation would require additional research.

5. Reliance on expert observation to evaluate controller performance.

Response: Expert observation of participant performance is a widely used research technique, and only one of many sources of information used in these simulations to evaluate elements of the Concept of Operations. Our approach was to collect objective data (system variables, communications data, clearance data, etc.), subject matter expert (SME) observations, and subjective responses from the participants themselves using questionnaires, online rating equipment, and caucus discussions. Observation of air traffic controller performance has been researched in our laboratory, and data collection forms for both terminal and en route controllers developed so that the right information is collected and by an instrument that has been used successfully many times. SME observations are part of nearly every simulation conducted here in human factors research, and in many other types of simulations.

<u>Conclusions</u>:

The Study report describes an experiment that was conducted to examine two research questions which do not adequately reflect the NYICC concept.

Response: The study reflects the aspects of the NYICC concept we were tasked to evaluate. The report clearly notes that we were not evaluating all aspects of the concept, all elements of NY airspace, etc. It is not practical to evaluate all elements simultaneously.

While the experiment and analysis was mostly sound, the results cannot be said to completely support the NYICC concept.

Response: Because we did not simulate the entire NYICC concept, we agree it does not "completely" support it. The data we did obtain are supportive of the concepts of collocation and use of terminal separation standards in en route airspace. We clearly qualified these findings because of the other components of the system not simulated, potential changes in the NY airspace, whether the collocated facility is well designed, etc.

The results provide support for a hypothesis that operations will be improved by collocating adjacent sector controllers.

Response: The data indicating support for collocation, including a few results that are not supportive, are based on the Collocated condition versus the Normal condition.

The confounding of the condition in which the separation standards were reduced makes this particular condition of the Study uninterpretable.

Response: We disagree that the results of the study are uninterpretable regarding reduced separation standards. While the study design does not allow us to determine whether or how much benefit could be obtained by using only terminal separation standards without collocation, comparisons of data in the Terminalized condition shows in general small improvements in comparison to the Collocated-only condition and larger improvements compared to the Normal condition. Use of terminal separation standards without collocation was not part of the NYICC *operational environment concept, and so was not simulated. If the concept were changed to that situation, then additional research would be required.*

In general, the NYICC concept may provide benefits for reducing congestion and delays in the Northeast. The Study overlooks potential changes in airspace and procedures that would have to occur in a real integration of ZNY and N90 which would potentially reduce or eliminate the conditions that this Study set out to investigate.

Response: It is completely impractical to simulate every aspect of the NYICC concept with every element of the NAS included. The Study addressed the elements of the concept tasked by the sponsors who needed empirical data to determine whether those aspects were viable. The sponsors were also collecting other data from other sources as input to the decision process. We used the sectors of interest to the Airspace Design team, presumably because they represent current problem areas, are representative of the larger problem, and/or are represented in other research efforts. Representatives of other organizations conducting fast-time modeling and cost-benefit studies observed both sets of simulations. The Study did not "overlook" other aspects of the concept; it simply limited what could be done to the primary concerns at the time. Furthermore, we clearly qualified our conclusions (see p. 103-104) regarding the potential impact of other changes on these findings.

<u>Questions:</u>

1. While the Study examined reduced losses of separation by extending the 3 nmi separation standard, it's not clear if this is improvement just because the separation standards were now less than 5 nmi, or if the controllers in the en route sectors were actually trying to meet the 3 nmi separation standard (which would have an impact on increasing capacity). In other words, did they actually space at 3 nmi, or did they just have fewer separation losses due to the reduced minimum standard?

   Response: We agree with the reviewer that if the participants were actually trying to meet the 3 nm separation standard, there would be an increase in capacity. We did observe increases in capacity and are willing to conclude, as does the reviewer, that the participants were using the 3 nm separation standard to some extent. In addition to fewer losses of separation with the reduced separation standard, the report also details other benefits. These other benefits were not likely to occur if the participants were not using the 3 nm separation standard to some degree. The en route participants reported on questionnaires they were better able to manage the traffic flow under reduced separation, and especially that they were better able to bring aircraft out of holding with 3-mile separation than with 5 mile. Our data reduction and analysis software is not capable of determining how close the controllers ran the aircraft under the Terminalized condition, but we believe they must have *been taking advantage of the reduced separation on the basis of the increased throughput and participant reports.*

2. It is unclear why the handoff position is considered a separate position from the radar position. Handoff position results should be combined with the radar position, or it should be analyzed completely separately. Handoff positions are not equivalent in responsibility or workload to the radar position. I would have liked to have seen some statistical analysis that showed how the handoff and radar positions produced effectively equal responses, which would have provided justification for combining their data.

Response: We conducted three basic types of analyses. Throughput variables that were a function of all participants in a scenario were not analyzed statistically. Other variables, such as the safety measures, were analyzed statistically using a Sector-by-Condition ANOVA. In these cases, both the radar and handoff positions contributed to the dependent measure, but we could not sort out the contributions separately by position. The remaining measures were analyzed using a Position-by-Condition ANOVA because of the capability for different levels of behavior or experience. These included intersector FTF communications, workload ratings, SME over-the-shoulder ratings, and participant questionnaire ratings. In those outcomes in which there were no significant differences between positions, then there were no differences between radar and handoff positions across the sectors simulated within the en route and/or terminal environments. Those data could be combined from a statistical perspective, but it would not add any information value in evaluating differences between conditions, our primary concern. That has already been factored into the analyses. That is, the statistical test of significance for the Condition variable is averaged over the input from the positions within each condition. On those variables (e.g., workload ratings) in which there were statistically significant differences between the radar and handoff positions, it is statistically improper to combine the data. It would also represent a loss of information and could result in misleading information about the primary variable of condition.

<u>Corrections:</u>

1. Section 3.2, the description of what TMA does is incorrect. TMA does not "assist en route controllers in selecting routes and time schedules to TRACON meter fixes to optimize arrival throughput." TMA creates schedules that are manifested in the form of metering list times and delay values, but does **not** provide any route advisories.

Response: We agree that the reviewer's definition of what TMA does and does not present to the controller is technically correct, and if desired by the sponsor we will change our description in the text. What we said about TMA is the summarization of the TMA description in the Free Flight Phase I Performance Metrics document (FAA, 1999) that describes how TMA uses radar *track data, flight plan data, and local meteorological conditions in its trajectory models "to compute routes and optimal schedules to the meter fixes for all arriving aircraft which have filed IFR flight plans, with consideration given to separation, airspace, and airport constraints." for the purpose of providing "ARTCC personnel with a means to optimize the arrival throughput of airports." (p. 2-1). That is, in our original text we were describing the process and purpose of TMA rather than the information displayed to the controller. We recommend we change our original text to explain what TMA does, what it displays, and what the controller does with it for the purpose of optimizing arrival throughput.*

Federal Aviation Administration. (1999). FFP1 performance metrics: An operational impact evaluation plan. Retrieved September 19, 2002, from http://ffp1.faa.gov/approach/media/FFP1metricsplan_v1_AOZ1.doc

2. Rognin & Blanquart's paper was published in 2000, not 2001.

Response: We completely agree, and we reported it as being published in 2000 in text (p. 3) and in the reference list (p. 106). We assume this proposed correction was an oversight by the reviewer.

Proposed Resolution: The authors believe that we have addressed nearly all the reviewer's concerns. We propose to add a statement about the increased possibilities of chance occurrence of results when reporting marginal findings in Section 3.0. We also propose adding a statement about the potential for chance findings when only presenting descriptive statistics (p. 18). Finally, we propose to reiterate the possibility of chance occurrences in the General Discussion and Conclusions section as a final caution. However, we also propose including using the overall pattern of results as an indication that the effects are systematic rather than chance occurrences.

In addition, if the sponsor desires, we will modify our description of TMA (p. 17) to more clearly indicate how the system works and what information is displayed to the controller.

We do not believe any other changes are required or warranted, other than technical editing for format, prior to publication.

Second NASA Review of "Effects of Collaboration and Reduced Lateral Separation Standards in the New York Integrated Control Complex"

This report describes a full mission human-in-the-loop simulation with air traffic controllers that was conducted to evaluate two changes proposed in the New York Integrated Control Complex concept of operations. One change is to collocate en route and TRACON operations to improve communications and coordination. The second change is to use terminal separation standards in one of the en route sectors (#74). A companion document - "New York Integrated Control Complex (NYICC): Concept of Operations" - describes many potential benefits of combining en route and TRACON operations in a single facility.

The human-in-the-loop study simulated operations in six sectors and ran nine traffic scenarios in a two day period. The traffic scenarios had traffic counts approximately 30% greater that current busy traffic conditions.

Overall the study and report seem to this reviewer to be a competent state-of-the-art simulator evaluation of two of the proposed procedural changes. The data show consistent, if not dramatic, improvements in system performance metrics mainly for the "Terminalized" condition in which the separation standards were reduced in one of the en route sectors. The report does a good job of presenting the descriptive data that supports the statistical results.

The following are specific comments and suggestions on improving the study and/or report.

4. A diagram showing the sector boundaries and nominal traffic flows would be helpful to readers not familiar with the airspace.

Response: Funding and schedule were not available to address this issue.

5. I think that study would have been stronger and shown a larger effect if the controllers had either been trained on or allowed practice time to develop procedures for working traffic with reduced separation standards.

Response: This is possible. However, the simulation schedule, which was restricted by participant availability and time/cost constraints, did not allow for such training.

6. The study used "frequency based" measures of sector performance. These measures seem to muddy the results because it is not possible to sort out if, for example, aircraft flew a longer distance in a sector or the handoff to the next sector occurred later. The use of these "frequency based" metrics results in statements such as - "the change may have been due to or due to ..." It seems that performance measures such as distance flown within a physical sector would eliminate this unnecessary ambiguity.

Response: We actually began doing some sector-based analyses. However, these analyses took considerable time and effort to complete because we had to define the FPAs for each sector and *take extras steps in the data reduction and analysis. Funding and schedule were not available to address this issue.*

7. Another ambiguous measure is the number of holds. It would seem that a metric of the total delay per aircraft that included holds, path vectors and/or speed reductions would be clearer.

Response: We are not able to derive such a metric at this time. Funding and schedule were not available to address this issue.

8. The controller TMA & URET tools are mentioned at the beginning of the section on performance metrics but it is not clear if these tools were actually used in the study (pg. 17).

Response: Clarify text to indicate that neither TMA nor URET was used in either experiment.

9. It was not clear what the definition of an operational error was in the terminalized Sector 74. The report just mentions the normal en route definitions of separation violations (pg. 32).

Response: The latest version of the text includes a more precise definition of operational errors and how they were calculated.

Final NASA Review and Recommendation

I have discussed the comments we had on the New York Integrated Control Concept report and the FAA responses with the NASA reviewers. The following is a summary of our recommendations:

The NASA reviewers have no fundamental concerns regarding the scientific merit of the study, given the scope and constraints under which it was performed. Our main interest is in limiting the interpretation and applications of the findings to what is reasonable, based on the data. The report indicates how the findings are to be interpreted, and places caveats where needed. The changes in the document suggested by the authors in response to our review are acceptable.

NASA strongly recommends that this study be taken in context. It addresses only some aspects of New York Integrated Control Concept. While it contributes to the knowledge on the identified issues, it should not be taken on its own as the pivotal factor in deciding to combine the facilities. Other studies of facility layout, airspace changes, procedures, safety, and cost/benefit are also needed to complete the evaluation of the concept.

Please contact me if you have any questions.