Advanced Controller Training In a Virtual Environment - 1

Nicole S. Racine, Titan Corporation
Edmundo A. Sierra, Jr., Titan Corporation

January 24, 2005

DOT/FAA/CT-TN05/4

Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

U.S. Department of Transportation
Federal Aviation Administration

NOTICE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/CT-TN05/4		

4. Title and Subtitle	5. Report Date
Advanced Controller Training in a Virtual Environment -1	January 2005
	6. Performing Organization Code ACB-330
7. Author(s) Nicole S. Racine and Edmundo A. Sierra, Jr., Titan Corporation	8. Performing Organization Report No. DOT/FAA/CT-TN05/4
9. Performing Organization Name and Address Federal Aviation Administration William J. Hughes Technical Center Atlantic City International Airport, NJ 08405	10. Work Unit No. (TRAIS)
	11. Contract or Grant No.
12. Sponsoring Agency Name and Address Federal Aviation Administration Office of Runway Safety and Operational Services 490 L'Enfant Plaza Suite 725 Washington, DC 20024	13. Type of Report and Period Covered Technical Note
	14. Sponsoring Agency Code Office of Runway Safety and Operational Services

15. Supplementary Notes

16. Abstract

This report describes the Anchorage ACTIVE-1 study which was a real-time, human-in-the-loop (HITL) simulation conducted in April 2004 by the FAA Alaska Region, William J. Hughes Technical Center (WJHTC), and Civil Aerospace Medical Institute (CAMI). The study's objective was to determine whether exposure to a series of operationally challenging scenarios in a state-of-the-art tower simulator would result in measurable performance improvement by fully qualified and current tower controllers.

The University of Alaska, Anchorage (UAA) tower simulator was used to emulate the Ted Stevens Anchorage International Airport (TSAIA) tower operations at high fidelity levels. Twelve air traffic controllers with experience from TSAIA participated over ten days of simulation. Traffic scenarios were created using TSAIA field data that was modified to create the desired traffic operations for the evaluation. The results showed significant improvements in detection of specific scripted errors, as well as general trends of improvement in other system performance measures. Participants identified some improvements needed to the simulator fidelity, but were generally positive about their experience. Overall, the results indicated that the tower simulator has great potential as a training tool.

17. Key Words	18. Distribution Statement
High-fidelity tower simulator, air traffic control, human error, human performance, ACTIVE-1	This document is available to the public through the National Technical Information Service, Springfield, Virginia, 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	77	

Form DOT F 1700.7 (8-72) | Reproduction of completed page authorized | Technical Report Documentation Page

Acknowledgements

This report reflects the contribution of many individuals from various organizations, and the authors would like to acknowledge and thank them for their efforts. Without them, this simulation would not have been a success.

We would like to thank Julia Pounds (FAA Civil Aerospace Medical Institute) for her assistance during the initial stages of the study and data analysis planning.

Les Habig, Boots Wilson, and Mark Tomlin provided Air Traffic Control subject matter expertise for scenario development, shakedown activities, and served as expert observers. Jim Krause served as an expert observer during testing.

Shaun O'Neil served as the Operator Station staff and was responsible for setting up and running the simulator, as well as scenario development activities. Mike Workman and Marcus Holmquist were part of the simulator technical support staff.

We would also like to thank the University of Alaska, Anchorage for the use of their facilities, staff, and pseudo-pilots throughout development and testing.

Table of Contents

Appendices

List of Illustrations

Figures Page

Tables Page

Executive Summary

The Federal Aviation Administration (FAA) has long recognized the value of simulator training. For pilots, air carriers and the FAA have both placed ever increasing levels of trust in the value of flight simulators as their handling qualities, visual systems and faithful replication of the flight experience have improved over the years. In fact, a pilot can be type certified without ever having flown the actual aircraft. Similarly, the FAA has utilized simulator training for en route air traffic controllers before fielding new equipment or implementing new procedures.

In general, simulator development for air traffic control tower (ATCT) applications has lagged in comparison to other air traffic domains and flight simulator development. Air traffic tower controllers, unlike their pilot counterparts, do not receive continuation training in simulators. Very few high quality tower simulators existed until recent advances in technology made available a number of high fidelity tower simulators.

One might assume that, if tower controllers were provided rigorous continuation training in advanced simulators analogous to the training provided to part 121 air carrier pilots, then their performance would be similarly enhanced. The FAA Office of Runway Safety and Operational Services sponsored the Advanced Controller Training in a Virtual Environment (ACTIVE-1) study to test this hypothesis and to examine the utility of providing runway incursion training to tower controllers in a high fidelity ATCT simulator. The goal of such training was to improve controller recognition and management of common factors leading to runway incursions (RI), and, when faced with unanticipated RIs, to rapidly take appropriate mitigation measures. ACTIVE-1 was intended to be the first in a series of simulations to address these issues.

This report describes the Anchorage ACTIVE-1 study which was a real-time, human-in-the-loop (HITL) simulation conducted in April 2004 by the FAA Alaska Region, William J. Hughes Technical Center (WJHTC), and Civil Aerospace Medical Institute (CAMI). The study's objective was to determine whether exposure to a series of operationally challenging scenarios in a state-of-the-art tower simulator would result in measurable performance improvement by fully qualified and current tower controllers.

The University of Alaska, Anchorage (UAA) tower simulator was used to emulate the Ted Stevens Anchorage International Airport (TSAIA) tower operations at high fidelity levels. Twelve air traffic controllers with experience from TSAIA participated over ten days of simulation. Traffic scenarios were created using TSAIA field data that was modified to create the desired traffic operations for the evaluation. The results showed significant improvements in detection of specific scripted errors, as well as general trends of improvement in other system performance measures. Participants identified some improvements needed to the simulator fidelity, but were generally positive about their experience. Overall, the results indicated that the tower simulator has great potential as a training tool.

Acronyms

AAL-1R	Alaska Region Runway Safety Office
ACTIVE-1	Advanced Controller Training in a Virtual Environment
ASDE	Airport Surface Detection Equipment
ATC	Air Traffic Control
ATCT	Air Traffic Control Tower
CAMI	Civil Aerospace Medical Institute
CAP	Critical Action Points
CPC	Certified Professional Controller
DBRITE	Digital Bright Radar Indicator Tower Equipment
E/O	Expert Observer
FAA	Federal Aviation Administration
HITL	Human-in-the-Loop
LOS	Loss of Arrival Separation
NAS	National Airspace System
NASA TLX	NASA Task Load Index
PI	Principal Investigator
RI	Runway Incursion
SME	Subject Matter Expert
SOP	Standard Operating Procedure
SOS	Study Operator Station
TRACON	Terminal Radar Approach Control
TSAIA	Ted Stevens Anchorage International Airport
UAA	University of Alaska, Anchorage
WJHTC	William J. Hughes Technical Center

1. Introduction

This report describes the study Anchorage Advanced Controller Training in a Virtual Environment (ACTIVE-1). ACTIVE-1 was a real-time, human-in-the-loop (HITL) simulation conducted in April 2004 by the Federal Aviation Administration (FAA) Alaska Region, William J. Hughes Technical Center (WJHTC), and Civil Aerospace Medical Institute (CAMI). The University of Alaska, Anchorage (UAA) tower simulator was used to emulate the Ted Stevens Anchorage International Airport (TSAIA) tower operations at high fidelity levels. Twelve air traffic controllers with experience from TSAIA participated over ten days of simulation. Traffic scenarios were created using TSAIA field data that was modified to create the desired traffic operations for the evaluation.

1.1 Background

The FAA Office of Runway Safety and Operational Services sponsored the ACTIVE-1 study to examine the utility of providing runway incursion training to tower controllers in a high fidelity air traffic control tower (ATCT) simulator. The goal of such training was to improve controller recognition and management of common factors leading to runway incursions (RI), and, when faced with unanticipated RIs, to rapidly take appropriate mitigation measures. ACTIVE-1 was intended to be the first in a series of simulations to address these issues.

Safe and efficient operation of the National Airspace System (NAS) rests on the performance of numerous professionals. Chief among the professional groups that impact safety and efficiency are air traffic controllers and pilots. Both groups are subject to rigorous qualification training, currency requirements, and periodic evaluations.

The FAA has long recognized the value of simulator training for pilots. Air carriers and the FAA have both placed ever increasing levels of trust in the value of flight simulators as their handling qualities, visual systems and faithful replication of the flight experience have improved over the years. In fact, a pilot can be type certified without ever having flown the actual aircraft.

Beyond initial training and certification, flight simulators play a significant role in air carrier continuation training programs. Pilots are afforded an opportunity to refresh systems knowledge, hone procedural and crew coordination skills and practice emergency procedures in a realistic but non-threatening environment. Although few pilots will face the catastrophic emergencies they routinely train for in the simulator, the few who do have been afforded the opportunity to rehearse their responses.

In general, simulator development for ATCT applications has lagged in comparison to other air traffic domains and flight simulator development. Very few high quality tower simulators existed until recently. Today, thanks to advances in computers and visual display systems, there are a number of high fidelity tower simulators available. Typical systems are capable of presenting a dynamic 360-degree field of view with a large number of aircraft, variable ceiling and visibility, and multiple communication channels controlled by electronic touch screens. Simulators can also be equipped with other systems common in the tower cab such as wind instruments, Digital Bright Radar Indicator Tower Equipment (DBRITE) scopes, approach control links, etc.

Air traffic tower controllers, unlike their pilot counterparts, do not receive continuation training in simulators. One might assume that, if tower controllers were provided rigorous continuation training in advanced simulators analogous to the training provided to part 121 air carrier pilots, then their performance would be similarly enhanced. However, this hypothesis remains largely untested. If found to be true, routine simulator training for tower controllers may help the FAA achieve operational error and runway incursion goals spelled out in the FAA Flight Plan (FAA, 2003).

1.2 Objective

The study's objective was to determine whether exposure to a series of operationally challenging scenarios in a state-of-the-art tower simulator would result in measurable performance improvement by fully qualified and current tower controllers. The objective is met if the selected performance measures show changes in skill and/or knowledge that can be used to determine whether successful performance was achieved.

1.3 Method Overview

In this study, twelve air traffic controllers with experience at TSAIA participated in this study. Controller teams consisted of a Local and a Ground controller. Controllers who were scheduled to participate for two days operated the Local control position on one day and the Ground control position on the other day. Controllers scheduled to participate for one day operated either the Local or the Ground position. Participants were instructed to control traffic as they would in the field throughout four scenarios. The scenarios included several planned critical action points (CAPs). Many CAPs had the potential to lead to runway incursions as determined from the examination of historical (within the last five years) common controller errors.

2. Empirical Evaluation

2.1 Organizational Roles and Responsibilities

2.1.1 Simulation Team

The Simulation Team consisted of the individuals developing and conducting the simulation. The team included the Principal Investigator (PI), human factors researchers, subject matter experts (SMEs), and simulator personnel.

The PI was responsible for the overall management of the simulation and was vested with the authority to direct the activities of all members of the Simulation Team as they relate to the simulation. The PI was a representative from the FAA WJHTC Simulation and Analysis Group, ACB-330.

Human factors researchers from FAA WJHTC and CAMI had specified roles in the simulation, administered forms, and conducted participant briefings and debriefings. These members of the team were continuously available in the test area to support the PI.

The Alaskan Region Runway Safety Office (AAL-1R) was the project manager. AAL-1R was responsible for developing the objectives, resource identification and acquisition, financial management and provision of pre- and post-simulation support activities.

The UAA tower facility staff provided simulation development resources and was responsible for simulator operators and pseudo-pilots.

2.1.2 Sponsoring Organization

The FAA Office of Runway Safety and Operational Services was the sponsoring organization of the ACTIVE simulation. The Office of Runway Safety and Operational Services funded the study and maintained overall authority for all aspects of the study.

2.2 Participants

Twelve current, Certified Professional Controllers (CPCs) from the TSAIA control tower volunteered for participation during the two weeks of simulation. Eight of the twelve participated twice, once as a Local controller, once as the Ground controller. The initial position, Local or Ground, was randomly determined. Though controllers participated twice, independence was assured by the switching of positions, counterbalancing of scenarios, and by having a minimum of six days between participation. The result was a sample size of ten (*N* = 10) for each position. The participants' average experience as tower controllers was 11.8 years (*SD* = 5.7 years).

Participation in this study was strictly voluntary, and the privacy of all participants will be protected. No individual names or identities were recorded or released in any reports. Each participant was assigned a participant code (e.g., L_1, G_1, L_2, G_2, etc.) that remained the same throughout the day of data collection. Those controllers that scheduled for two days of data collection were assigned a different code on their second day of participation (e.g., L_1, and G_6-L_1). Strict adherence to all federal, union, and ethical guidelines was maintained throughout the study.

Participants were subjected to minimal risk. Per the definition of minimal risk, the probabilities of harm or discomfort in this simulation were not greater than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests.

2.3 Apparatus and Facility Support

2.3.1 UAA Simulator

The UAA tower simulator is a full-scale ATCT simulator that provided an interactive, highly realistic environment for controllers. An image generator provided a full 360-degree out-the-window scene including moving aircraft, ground vehicles, and weather effects (see Figure 1). Displays that are similar to the DBRITE scope and Airport Surface Detection Equipment (ASDE) display were provided in the simulation.

The tower cab supported up to four simultaneous positions that may be configured for Local controllers and Ground controllers, flight data/clearance delivery, traffic management coordinator, tower cab coordinator, or supervisors. Available support positions included ramp controllers, pseudo-pilots, Terminal Radar Approach Control (TRACON) controllers, and airport operators. The pseudo-pilots used a graphical interface to view the airport and surrounding airspace and for entering aircraft control instructions.

All simulation participants used a digital voice communications system (e.g., radios, headsets, telephones and interphones) to interact in real-time. The tower simulator recorded for play back all data elements available in the air traffic control simulation. Surface movement metrics, such as taxi times and runway crossings, were collected and reported for each run. Video monitoring of activities in the tower cab during simulations were also recorded for post simulation playback and analysis.

Figure 1. TSAIA displayed on UAA tower simulator.

2.3.2 Airport

This simulation emulated TSAIA (see Appendix A for airport diagram). It included the control tower and the entire airport. Anchorage, Alaska was chosen as the study site due to the collocation of several key resources including a high-fidelity tower simulator located at UAA, the availability of tower controllers from a medium sized ATCT (i.e., TSAIA), and a Regional Runway Safety Office to coordinate the effort.

2.3.3 Pseudo-Pilots

Six trained pseudo-pilots supported the operation by emulating all pilot actions. UAA trained these individuals to an acceptable level of operational proficiency, familiarity with the airport, and its standard operating procedures. Additional UAA pseudo-pilots and TSAIA personnel were available to handle the communications between pilots and participant controllers.

2.3.4 Study Operator Station Staff

The Study Operator Station (SOS) staff was responsible for setting up the simulation environment (e.g., starting each run, selecting the proper scenario, setting the weather conditions, aircraft start positions, etc.). This individual was a UAA staff member familiar with the simulator, the airport, and its standard operating procedures.

2.3.5 Expert Observer

Two Expert Observers (E/Os) from TSAIA sat in the tower cab. One E/O observed the Local control position and the other observed the Ground control position. The E/Os were SMEs with ATCT experience, were familiar with the airport, and were familiar with its standard operating procedures. The E/Os recorded data on the E/O Over-the-Shoulder Form (see Appendix B). The E/Os also participated in the Post-Run debriefings.

2.4 Design

This study of the effect of practice in a simulated ATC Tower on performance used a repeated-measures design. *Practice* was the manipulated variable of interest. Dependent measures are listed in Table 1 and Appendix C. We collected measures of planned hear-back/read-back error detection, planned conflict detection, and total system effectiveness measures during each scenario. We compared these measures to each other to test the effect of Practice on performance in each of these areas. The reader should note that individual controller performance was not under evaluation. We used the performance measures to determine whether the ATC Tower simulator is an effective training tool.

We exposed controllers to a series of four scenarios in the simulator. These four scenarios (scenario 1 [pretest], scenario 2, scenario 3, and scenario 4 [posttest]) were the levels of Practice. Scenarios 1 and 4 were the principal scenarios of interest. Controllers worked the first of the scenarios without instruction or knowledge of the nature of the conflicts. We took measures during this scenario to use as the baseline, or pretest. Scenarios 2 and 3 were followed by knowledge of results, a review of warning signs, and mitigation strategies. We used the last scenario, scenario 4, as a final measure of performance, or posttest. We compared the measures for every scenario to each other; however, scenarios 1 and 4 were compared to each other and used to determine whether changes in performance resulted from practice.

2.5 Procedures

2.5.1 Instruction to Control Teams

We instructed participants to perform as they normally would in the field using the same consideration for efficiency and vigilance for safety that they would in a real-world environment. Participants were not given information about any off-nominal scenario characteristics during the scenario. We also did not give participants feedback regarding whether a scripted event occurred, unless they elicited feedback about an event as they would during actual operations or the scenario had ended and they were debriefing with the E/Os.

2.5.2 Familiarization

At the beginning of each day, controllers experienced a twenty-minute familiarization run to acclimate them to the simulator environment.

2.5.3 Scenarios

We used four data collection scenarios and one familiarization scenario for the simulation. The familiarization scenario was approximately 20 minutes in length, and provided the participants with the opportunity to become familiar with the simulator environment. Each of the data collection scenarios was approximately 45 minutes in length, and included a variety of scripted CAPs that could lead to runway incursions unless the CPC took appropriate action (see Appendix D for example descriptions of scripted events).

Based on a five year analysis of the Database Management for Runway Safety, maintained by the FAA Office of Runway Safety and Operational Services, the most common controller errors associated with runway incursions are as follows: loss of arrival/departure separation, aircraft cleared to land/depart while another aircraft is cleared to cross the same runway, read-back/hear-back errors, and issuing instructions based on mistaken aircraft or vehicle locations. The same database shows that over 70% of all pilot runway incursions occur after a pilot receives and acknowledges a hold short instruction but then enters the runway environment anyway. Potential for these types of errors were scripted into the scenarios as CAPs.

The CAP can be thought of as the first event in a sequence of events leading to a runway incursion. A controller that recognizes developing problems, such as an impending loss of separation or a read-back error, responds in a manner that terminates the sequence of events (Objective Controller Response). If the controller did not provide an objective response at the CAP, then the pseudo-pilot sustained the runway incursion sequence.

Each scenario also contained instances of a pilot entering the runway without authorization. In these cases, the objective controller response mitigated the severity of the incursion. For example, a controller employing proper runway scan technique observed an aircraft crossing the departure end of the runway and canceled the departure's takeoff clearance before brake release.

The objectives of the simulation were partially satisfied by evaluating controller reaction to scripted events, such as their response to each scenario CAP; however, other measures were obtained to support the conclusions that we drew.

All scenarios utilized identical environmental conditions chosen to optimize the controllers' ability to visually acquire aircraft and vehicles in the airport environment. The intent was to minimize performance variability due to differences between simulator and real world acuity.

Traffic characteristics were modeled by taking actual TSAIA traffic samples from August 2003. Traffic levels approximated 75% of actual levels so that the scenario was manageable and realistic for a two-person controller team.

2.5.4 <u>Data Collection</u>

We collected subjective and objective data throughout the study. Table 1 lists the data collection instruments that we used. Performance data collection is listed in Appendix C and is broken down by Safety, Efficiency, and Capacity constructs. In addition, measures for assessing training are broken down into the CAP error type categories.

Table 1. Real-Time Data Collection Instruments

Instrument	**Objective**
Simulator Data Recording	To collect aircraft performance data relating to airport operations such as arrival and departure statistics. Provides replay capability for post hoc analysis.
Audio Recordings	For communication analyses, post simulation replay, and to provide backup source for data.
Video Recordings	For post simulation replay and to provide backup source for data
E/O Over-the-Shoulder Form	Gather information regarding simulation events

Each run was video and audio recorded to capture the activity in the control tower. We used the recordings to gather supplemental data and to substantiate other subjective and objective data.[a]. Audio recordings captured the ambient conversations between controllers and the transmissions on all simulated frequencies. Video recordings captured general views of the tower control positions and the out-the-window view of interest.

We collected subjective data from participants using questionnaires, situation awareness ratings, NASA Task Load Index (NASA TLX) (see Appendix E), and debriefing sessions. In addition, two E/Os collected observer data using an over-the-shoulder rating form for each position. Table

[a] Video and audio recordings were used exclusively for the reasons stated. Participants signed a release giving their permission to record and use this information as stated.

2 summarizes the subjective data collection methods.

Table 2. Subjective Data Collection Methods

Instrument	Source	Frequency	Objective
Background Questionnaire	Participants	Once	Gather participant demographic information
E/O Over-the-Shoulder Form	E/O	Every run	To record real-time event data
Post-Run Questionnaire	Participants	Every run	Gather participant feedback concerning operations, events, and observations for the preceding run.
NASA TLX	Participants	Every run	Gather information related to controller perceived workload.
Post-Run Structured Debriefing	Participants	Every run	Researchers will question participants regarding unusual/unforeseen events during the run (e.g. scripted errors such as runway incursions).
Post-Simulation Questionnaire	Participants	Once	Gather information regarding benefits of the training sessions, controller ability to mitigate scripted errors, simulation fidelity, controller recognition of scenario similarities, etc.
Post-Simulation Debriefing	Participants	Once	Gather information that was not previously acquired

2.5.4.1 Expert Observer Over-the-Shoulder Forms

The E/Os collected data about events that occurred during each scenario. The E/Os took note of special events that occurred during a run, comments made by the controllers, and any equipment malfunctions. The E/Os completed an over-the-shoulder form for each scenario, including an overall rating form and documentation of controller actions related to the CAPs.

2.5.4.2 Questionnaires and Debriefings

During the initial briefing session, participants completed a Consent Form (see Appendix E) and a Background Questionnaire (see Appendix F). The Background Questionnaire's purpose was to obtain demographic information about our sample and solicit information related to pilot experience.

At the end of each run, the E/O and a human factors engineer conducted a short debriefing session and asked the participants to complete a Post-Run Questionnaire (Appendix G). The Post-Run Questionnaire solicited participant responses regarding information about the occurrence of unforeseen events (namely, to draw out information about scripted events).

At the end of all runs, participants completed a Post-Simulation Questionnaire (see Appendix H), followed by an overall simulation debriefing. The Post-Simulation Questionnaire solicited participant responses regarding information such as simulation fidelity, adequacy of familiarization training for simulation, and the usefulness of the training sessions for real world application.

All questionnaires contained space to provide additional information as appropriate.

2.5.4.3 NASA Task Load Index

A human factors engineer administered the NASA TLX (Hart & Staveland, 1988) after every run (see Appendix I). The NASA TLX was composed of six factors: mental demand, physical demand, temporal demand, performance, effort, and frustration level. As designed, NASA TLX has two parts. The first part asks participants to choose between one of a pair of factors that they feel caused them the most workload (e.g., mental vs. physical; physical vs. temporal). The second part of the NASA TLX asks participants to rate the level of workload they felt by placing marking an unnumbered scale. Only the second part was administered for this study. The ratings provided during pre- and post-test were not significantly different from each other.

3. Simulation Assumptions and Limitations

Though this particular study emulated aspects of the operational environment at high fidelity levels, all simulation research presumes some limitations and assumptions. It was beyond the scope of this study to compare tower simulation training with current training methods. Objectives for this study were considered to be met if there was an increase in human performance due to practice and training in the simulator.

Reactions to incursions may have been better in the simulation than they would be in the real world. Controller teams were particularly vigilant during the runs, which was expected to improve performance and situation awareness compared to actual control tower operations.

Due to some unforeseen technical problems, some unusual simulator behavior occurred. We recorded all anomalies, evaluated the impact or potential impact on results, and made appropriate adjustments in the analysis and data reported here.

We identified some equipment limitations prior to the commencement of data collection. Embedded within the first study objective was to determine if the tower simulator had adequate fidelity to achieve measurable controller performance improvement. Therefore, we collected subjective feedback as well as objective data to determine areas of needed simulator improvement, which are reported in the Results Section 4.3 and Discussion Section 5.

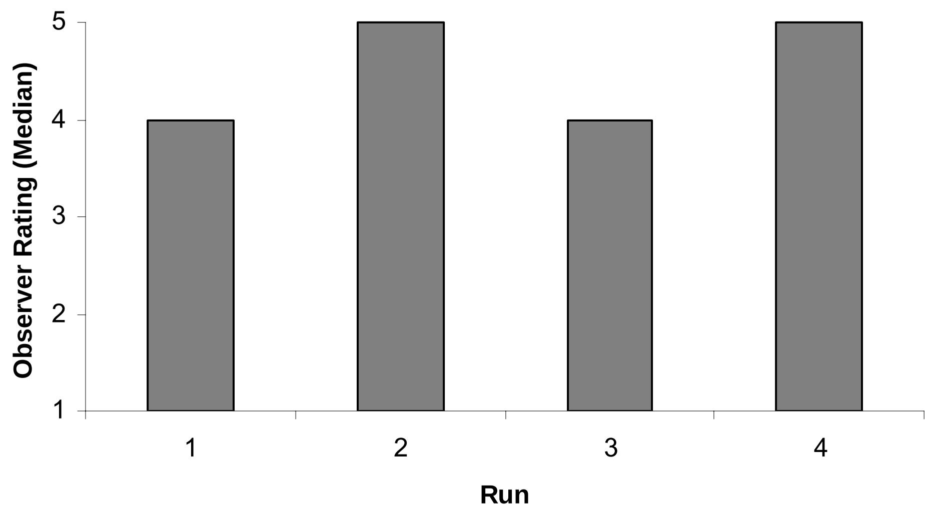
4. Results

4.1 Expert Observer Data

4.1.1 Over-the-Shoulder Form

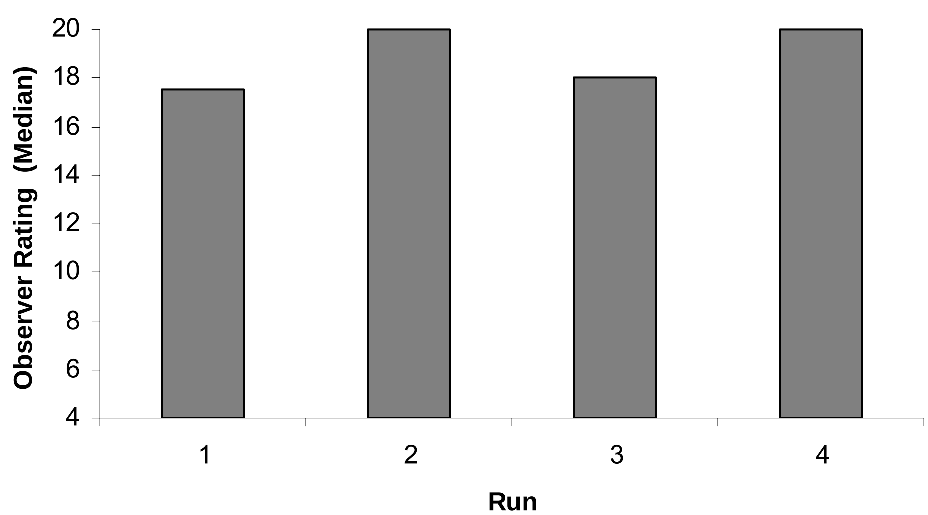
E/Os rated controller performance for each of the runs. The observers rated controllers on separation, coordination, control judgment, methods and procedures, equipment, and communication tasks. Control judgment, methods and procedures, and communication tasks were multidimensional; that is, there were a number of subtasks used to get a rating for these tasks (see Appendix B). The items on the form were worded as controller behavior (e.g., "Separation is insured") to which the observer assigned a rating of never, seldom, sometimes, often, and always.

Ground controller performance ratings improved significantly with experience in the simulator. We tested the hypothesis that controller performance improved with experience in the simulator as measured by subjective observer ratings. The null hypothesis was that subjective measures of controller performance did not increase with experience. The alternative hypothesis was that subjective measures increased with experience. A *Friedman two-way analysis of variance by ranks* (Friedman Test) showed statistically significant improvements in Ground controller separation ($F_r(3) = 15$, $p < .05$) and control judgment ($F_r(3) = 15.65$, $p < .05$) ratings. Multiple comparisons indicated that, for both separation (see Figure 2) and control judgment performance (see Figure 3), the difference between ratings given in the first and fourth run were statistically significant at an alpha level of .05. Local controllers did not show any improvement as measured by observer ratings of this type.



Note. The median does not capture that four ratings of 4 were assigned on Run 2, while only one 4 rating was given on Run 4.

Figure 2. Median observer ratings of ground controller separation.



Note. The median does not capture that four ratings less than 20 were assigned on Run 2, while all ratings on Run 4 were 20s.

Figure 3. Median observer ratings of ground control judgment.

4.1.2 Critical Action Points

4.1.2.1 Catching Hear-back/Read-back Errors

4.1.2.1.1 Objective Data

The E/O was asked to select one of the following four items that would best describe a controller's reaction to a pilot failure to read back hold short instructions.

1. Completely unobserved
2. Observed, but no action taken
3. Action taken at 1st opportunity (Succeeded? Failed?)
4. Action taken at 2nd opportunity (Succeeded? Failed?)

Successful responses (3 and 4) were expected to be the most frequent. Therefore, we were interested in whether success detecting a hear-back/read-back error increased with controllers experience in the simulator. For each participant, a ratio of successful detection of a hear-back/read-back error to those available (usually 4 per scenario) was calculated.

Statistical analyses indicated that experience resulted in improvement catching hear-back/read-back errors in the simulator. The null hypothesis tested was that experiencing scenarios in the simulator had no differential effect on catching the errors. The alternative hypothesis was that experiencing scenarios had an effect. The hypotheses were tested with the Friedman Test. Two participants were excluded from the analysis because one of the scenarios in their set was invalid due to random error. The results indicated that experiencing scenarios had a significant effect on detecting errors, $F_r(3) = 8.35$, $p < .05$. Multiple comparisons between conditions indicated that the difference between the first and last scenario was statistically significant at $p < .05$. Table 3 shows that six participants in the first scenario missed failures to read back hold short instructions; by the last scenario, only one participant missed a failure to read back instructions.

Table 3. Ground Control Accuracy Detecting Hear-Back/Read-Back Errors in Percentage

Ground Controller	Run 1	Run 2	Run 3	Run 4
G1	100	75	25	75
G2	75	100	100	100
G3	75	100	100	100
G4	50	75	75	100
G4	75	100	100	100
G6	100	50	100	100
G9	33	100	75	100
G10	50	100	77	100

Similar errors were planned for the Local position, but an analysis could not be conducted. Many planned errors could not take place because controllers gave unpredictable instructions to the subject aircraft preventing the scripted error. For example, some aircraft that were supposed to fail to read-back hold short instructions were instructed to go-around and could not land in time to make the error before the scenario ended. In other cases, the aircraft were not held short of the runway, but instructed to cross and contact ground. Because of the variable instructions, the amount of missed data prevented a valid analysis of improvement or decline in performance. Therefore, the hear-back/read-back analyses is only reported for the Ground position.

4.1.2.1.2 Observer Ratings Data

The E/O was asked to rate controller responses to events using a paper form (see Appendix B). Each form had six items that observers rated: recognition time, appropriateness of action, appropriateness of timing, accuracy of communication with traffic, accuracy of communication with ATC, and an overall rating of the response to the event. An example of the scales used to rate the items is depicted in Figure 4.

1	2	3	4	5
Very inadequate	Slightly inadequate	Borderline	Slightly adequate	Very adequate

Figure 4. Example of the rating scale used by expert observers.

Only differences in the overall ratings between scenarios were analyzed. Observers circled the same number for every item when they rated a scenario. For example, if the observer considered a response by the controller to be slightly inadequate, a 2 would be circled for all items. Therefore, only the overall rating was used for the analysis.

Statistical analyses indicated that experience resulted in improved observer ratings of handling of the event. The null hypothesis tested was that experiencing scenarios in the simulator had no differential effect on observer ratings of the adequacy of the response to events. The alternative hypothesis was that experiencing scenarios had a positive effect on the ratings. The hypotheses were tested with the Friedman Test. Two participants were excluded from the analysis because they had missing data in their set. The results indicated that experiencing scenarios had a significant effect on observer ratings of controller responses to the events, $F_r\,(3) = 8.48$, $p < .05$. Because of the level of significance we chose for the multiple comparisons, we were unable to conclude that the fourth run was rated significantly higher than the first; however, it was the largest difference between runs. Table 4 shows the median of the ratings given by observers. Note the difference in median ratings and variability of the ratings as the controllers experienced scenarios.

Table 4. Median Observer Ratings of the Overall Adequacy of the Controller's Response to the Event

		Run			
		1	2	3	4
Ground Controller	G1	4	3	1	5
	G2	1	5	5	5
	G3	3	5	5	5
	G4	3	5	5	5
	G5	5	5	5	5
	G6	5	3	5	5
	G9	1	5	5	5
	G10	3	5	5	5

4.1.2.2 Handling Loss of Arrival Separation

4.1.2.2.1 Objective Data

The E/O was asked to select one of the following four items that would best describe a controller's response to a loss of arrival separation.

1. Completely unobserved
2. Observed, but no action taken
3. Action taken at 1st opportunity (Succeeded? Failed?)
4. Action taken at 2nd opportunity (Succeeded? Failed?)

We used failures of any kind in the analysis of responses to loss of arrival separation. The task of correcting a hear-back/read-back error was relatively simple once it was caught. However, correcting a loss of arrival separation was a more demanding task. Therefore, ratios of any kind of failure [ratings of 1, 2, 3 (failed), and 4 (failed)] to planned losses of arrival separation were calculated. The ratio (usually $f_{\text{ratings}}/5$) was then compared between runs.

Controllers did not improve at correcting losses of arrival separation. The null hypothesis in the analysis was that experience had no differential effects on how often they failed to effectively deal with a loss of arrival separation. The alternative hypothesis was that experience had an effect. We used the Friedman Test for this analysis. We excluded two participants from the analysis because of missing data. We did not find statistically significant differences between conditions, $F_r(3) = 4.48$, $p > .05$.

We suspect that the incidence of responses 3 (with failure) and 4 (with failure) were more a consequence of simulator capability than inadequate controller responses to losses of arrival separation. Specific simulator limitations are discussed in section 5. If we excluded responses 3 and 4 and only ratings 1 and 2 are used, then there is insufficient data to conduct a statistical analysis. Those ratings are given with a frequency of 1, 1, 2, and 0 for runs 1 to 4 respectively. Consequently, we cannot come to a valid conclusion.

4.1.2.2.2 Observer Ratings Data

In addition, there were no improvements in observer ratings of the overall adequacy of the controllers' responses to the losses of arrival separation. We asked the expert observers to provide an overall rating of controller responses to the events. We then calculated the median of the ratings (for about five events per scenario) and analyzed them for differences resulting from experience in the simulator. The null hypothesis tested was that experience in the simulator did not result in different medians between any two groups. The alternative hypothesis was that medians were significantly different between two groups. No statistically significant differences were found, $F_r(3) = 2.86$, $p > .05$. These results were consistent with the objective observations of *failures* discussed in the first three paragraphs of Section 4.1.

4.2 System Performance Measures

4.2.1 Efficiency

We hypothesized that experience in the simulator would result in measurable improvement in the efficiency of the system. To test the hypothesis, we recorded several measures that we considered captured Efficiency (see Appendix C) and evaluated the statistical significance of the difference between the first scenario and the last. We used the first scenario as a pretest and the last scenario as a posttest. Though there was a trend showing improvement in the majority of the efficiency measures, the differences were not statistically significant.

4.2.1.1 Time Interval Between Landing Aircraft

We tested whether experience in the simulator had any effect on the time interval between landing aircraft. The simulator recorded the time that each aircraft landed on runways 6L and 6R. We conducted the analysis for each runway separately because the controllers use the runways differently (i.e., more 6L departures likely affected the number of arrivals).

To test the effect of experience in the simulator on the time interval between landing aircraft, we measured twice for each controller team. Controllers experienced four scenarios in the simulator. We measured the time intervals on the first scenario and the last scenario. The first scenario was the pretest and the last was the posttest.

We used a *permutation test for paired replicates* (hereafter, permutation test; Siegel & Castellan, 1988) to evaluate the statistical hypotheses. We chose the permutation test because of its power and because we could test the hypotheses without making any assumptions about normality or homogeneity of variance. The null hypothesis was that the pretest and posttest scenario time interval between landing aircraft was equivalent. The alternative hypothesis was that the time intervals were different for each scenario. We used seven pairs for the test (one set was lost because the data collection function was not turned on during the scenario; scenario durations were not equivalent for the other two sets). For runway 6L, the test showed that the observed sum of the differences (Σd_i = [242]) was not in the rejection region (alpha less than .05); therefore, we could not reject the null hypothesis. The same is true for runway 6R, Σd_i (-63). Figure 5 shows that the time interval between landing aircraft was shorter during the posttest for 6L, but shows an increase with experience for 6R.

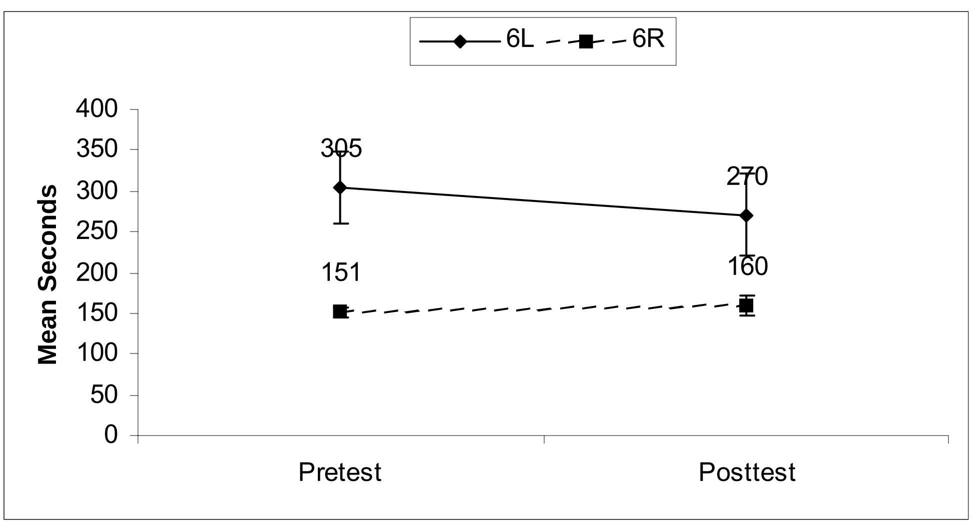


Figure 5. Mean (+/- standard deviation) time (seconds) interval between landing aircraft as a function of experience.

4.2.1.2 Time Interval Between Departing Aircraft

We tested whether experience in the simulator had any effect on the time interval between departing aircraft. The simulator recorded the time that each aircraft departed from runways 32 and 6L. We conducted the analysis for each runway separately because the controllers used runway 32 for departing aircraft more frequently. In addition, runway 32 was generally used for the heavy and larger aircraft, while 6L was typically used to depart smaller aircraft.

We used the permutation test to evaluate the statistical hypotheses. The null hypothesis was that the pretest and posttest scenario time interval between departing aircraft was equivalent. The alternative hypothesis was that the time intervals were different for each scenario. We used seven pairs for the test. For runway 32, the test showed that the observed Σd_i (75) was not in the rejection region; therefore, we could not reject the null hypothesis. For runway 6L, the test also showed that the observed Σd_i (85) did not reach statistical significance. Figure 6 shows a decrease in the time interval between landing aircraft on both runways. It also shows a high degree of variability for runway 32.

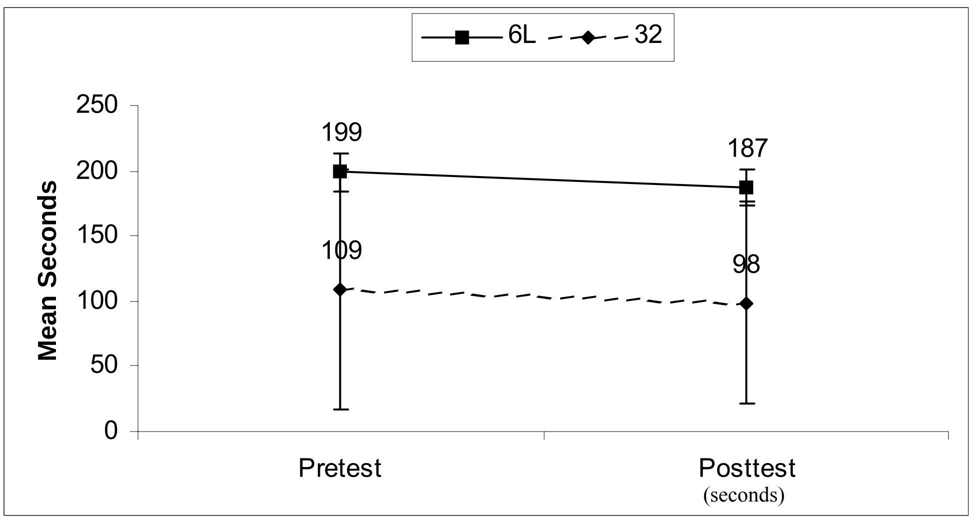


Figure 6. Mean (+/- standard deviation) time (seconds) interval between departing aircraft as a function of experience.

4.2.1.3 Number of Missed Approaches

We used a permutation test to evaluate the statistical hypotheses. The null hypothesis was that the pretest and posttest scenario number of missed approaches were equivalent. The alternative hypothesis was that the number of missed approaches was different for each scenario. We used seven pairs for the test. The test showed that the observed Σd_i (-3) was not in the rejection region; therefore, we could not reject the null hypothesis.

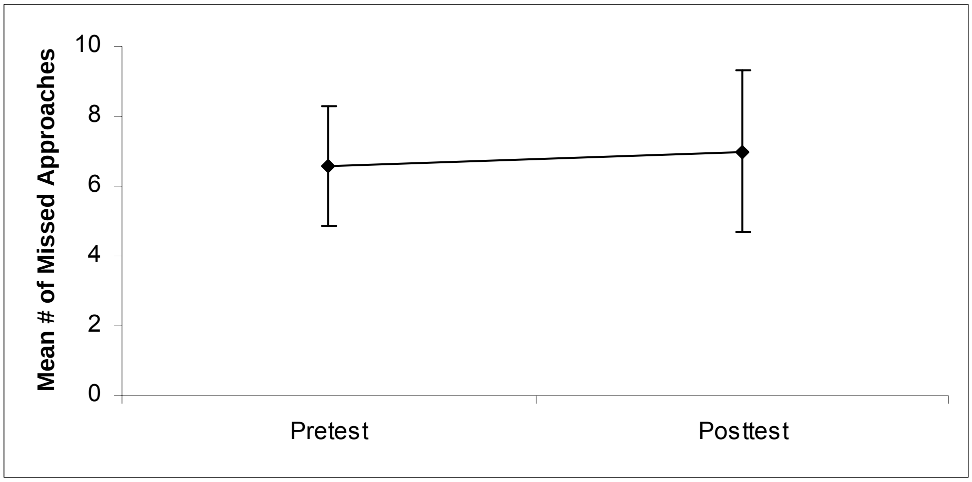


Figure 7. Mean (+/- standard deviation) number of missed approaches as a function of experience.

4.2.1.4 Taxi Time

We used the permutation test to evaluate the statistical hypotheses. The null hypothesis was that the pretest and posttest scenario taxi times were equivalent. The alternative hypothesis was that the taxi times were different for each scenario. We used seven pairs for the test. For taxi in time, the test showed that the observed Σd_i (-58.67) was not in the rejection region; therefore, we could not reject the null hypothesis. For taxi time out, the test also showed that the observed Σd_i (126) did not reach statistical significance. Figure 8 shows the average number of seconds for the two scenarios; we did not include the standard deviations for the taxi in (pretest *SD* = 30, posttest *SD* = 50) and taxi out (pretest *SD* = 96, posttest *SD* = 85) variability in the figure because they would not be discernable.

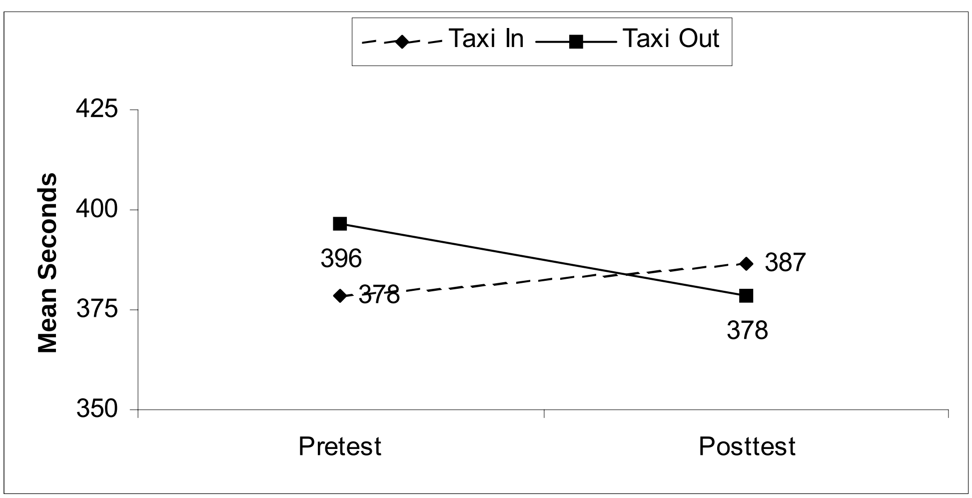


Figure 8. Mean taxi time (seconds) taxiing in and out as a function of experience.

4.2.1.5 Duration of Holds

We used the permutation test to evaluate the statistical hypotheses. The null hypothesis was that the pretest and posttest duration of holds on runways were equivalent. The alternative hypothesis was that the duration of holds was different for each scenario. We used seven pairs for the test. The test showed that the observed Σd_i (19) was not in the rejection region; therefore, we could not reject the null hypothesis.

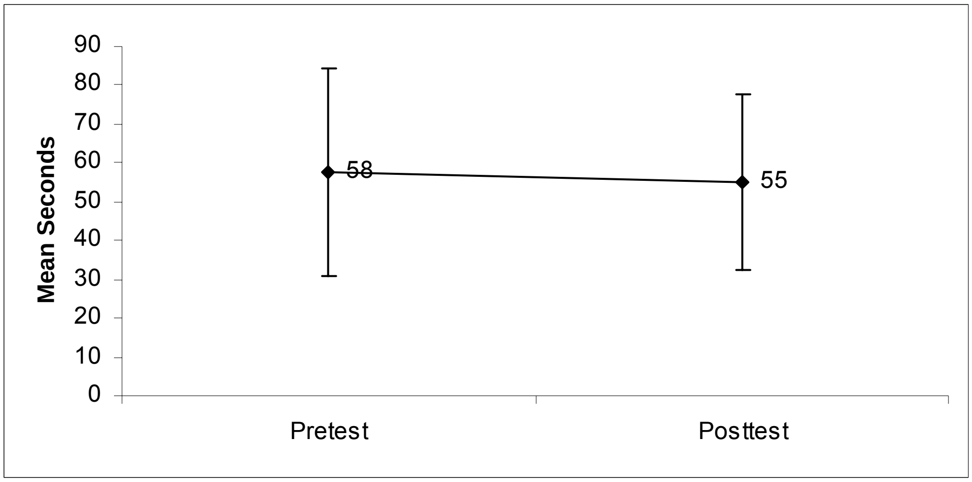


Figure 9. Mean time (seconds) of aircrafts holding on runways as a function of experience.

4.2.1.6 Other Efficiency Measures

Other Efficiency measures were also not significant. We considered the Efficiency measures listed in section 4.2.1.1 to 4.2.1.5 as the primary indicators of system efficiency. However, other measures may also show performance differences, especially if a number of them are different. Table 5 lists other analyses conducted using the permutation tests along with means and standard deviations. The results confirmed that there was no difference in Efficiency between the pretest and posttest scenarios.

Table 5. Other Efficiency Measures with Permutation Test Results, Means, and Standard Deviations for Pre- and Posttest Scenarios ($N = 7$)

Measure	Observed Σd_i	Mean		Standard Deviation	
		Pretest	Posttest	Pretest	Posttest
Average arrival rate 6L	17.17*	11.06	13.51	.99	2.08
Average arrival rate 6R	-8.94	23.09	21.81	1.05	1.29
Average departure rate 6L	-9.39*	13.10	11.76	1.84	2.06
Average departure rate 32	-.92	23.24	23.11	1.33	1.50
Average inbound stop durations (s)	-91.66	56.51	69.60	15.18	30.35
Average outbound stop duration (s)	-2.53	103.68	104.04	46.22	39.62
Average runway occupancy durations (s)	-19.95	70.07	72.92	2.95	8.02
Departure runway average runway occupancy durations (s)	125.17	110.75	92.87	26.01	25
Total non-movement area push-backs	0	26.71	26.71	.76	.76
Total outbound stop durations (s)	-713	2891.43	2993.29	1292.91	1770.35
Total outbound stops	5	28.43	27.71	3.99	6.10
Total runway exit count	13	25.86	27.71	1.57	.95

* $p < .05$

4.2.2 Capacity

The research team selected four measures to represent airport Capacity for each run. The first was throughput. We defined throughput as the number of airplanes that touched down and arrived at their terminal plus the number of planes that pushed-back then took off during the run. The second measure was arrivals. We defined arrivals as the number of aircraft that landed during the scenario. We also measured departures, defined simply as the number of airplanes that took off during a scenario. Finally, we used the number of aircraft handled throughout the scenario as the fourth measure. We took the four measures on runs 1 through 4, but used run 1 as a pretest and run 4 as a posttest for the analyses. Results of the analyses of each Capacity measure follow.

4.2.2.1 Throughput

We tested whether experience in the simulator had any effect on controller teams' throughput of aircraft. The simulator recorded the number of aircraft that pushed back from the gate and took off. It also recorded the number of aircraft that touched down and arrived at a gate. We added those two measures together to obtain the value for throughput.

To test the effect of experience in the simulator on throughput, we measured throughput twice for each controller team. Controllers experienced four scenarios in the simulator. We measured throughput on the first scenario and the last scenario. The first scenario was the pretest and the last was the posttest.

We used a permutation test to evaluate the statistical hypotheses. The null hypothesis was that the pretest and posttest scenario throughput was equivalent. The alternative hypothesis was that the throughput of aircraft was different for each scenario. We used seven pairs for the test (one set was lost because the data collection function was not turned on during the scenarios; the other two had pretest scenarios that ran too short). The test showed that the observed sum of the differences (Σd_i = 16) was not in the rejection region; therefore, we could not reject the null hypothesis. Visual inspection of the data shows that controllers' throughput was higher during the posttest (See Figure 10). The reader may also note that controller performance was also more consistent (i.e. less variable) after having some experience.

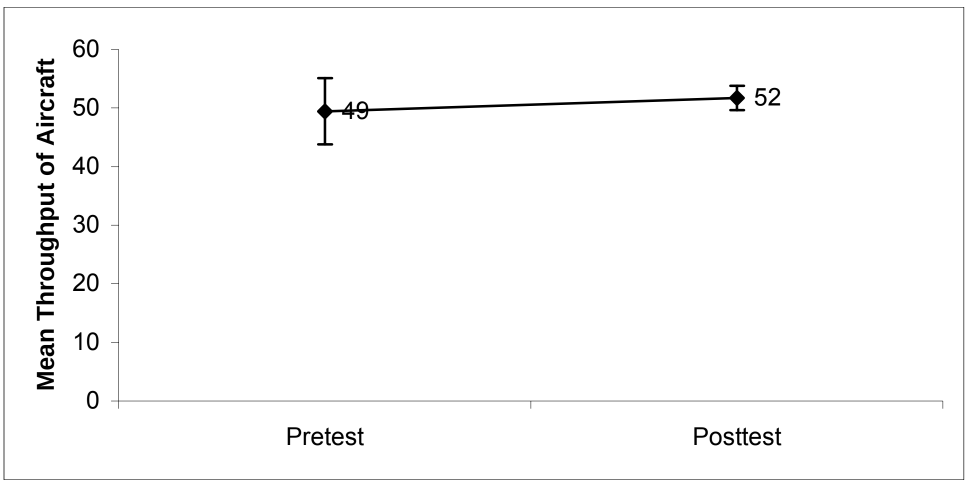


Figure 10. Mean (+/- standard deviation) throughput of aircraft as a function of experience.

4.2.2.2 <u>Arrivals</u>

We tested whether experience in the simulator had any effect on the number of arrivals in a scenario. We used only the number of aircraft that touched down during a run for this analysis. As in previous analyses, we used the measures taken during the first and last run, and used the permutation test to evaluate the hypotheses. The null hypothesis was that the numbers of pretest and posttest arrivals were equivalent. The alternative hypothesis was that the number of arrivals was different. We used seven pairs for the test (one set was lost because the data collection function was not turned on during the scenarios and one of the runs for the other two pairs was of a different duration). The test showed that the observed Σd_i (14) was not in the rejection region; therefore, we could not reject the null hypothesis that the number of arrivals was the same. Though not statistically significant, visual inspection of Figure 11 shows that the number of arrivals was greater during the posttest (See Figure 11; pretest *SD* = 1.77, posttest *SD* = 1.21).

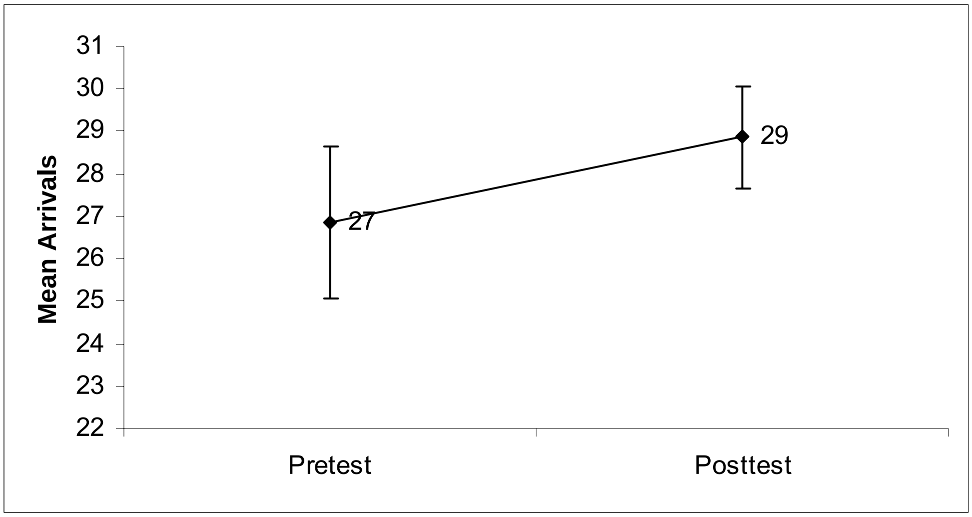


Figure 11. Mean (+/- standard deviation) number of arrivals as a function of experience.

4.2.2.3 Departures

We tested whether experience in the simulator had any effect on the number of departures in a scenario. Similar to previous analyses, we used the number of aircraft that took off during the first and last run. The null hypothesis was that the numbers of pretest and posttest departures were equivalent. The alternative hypothesis was that the number of departures was different. We used seven pairs for the permutation test. The test showed that the observed Σd_i (3) did not reach statistical significance; therefore, we could not reject the null hypothesis that the number of departures was the same. Note that Figure 12 shows the same pattern as the other Capacity measures of an increase in central tendency and decrease in variability (pretest *SD* = 2.94, posttest *SD* = .90) from pretest to posttest.

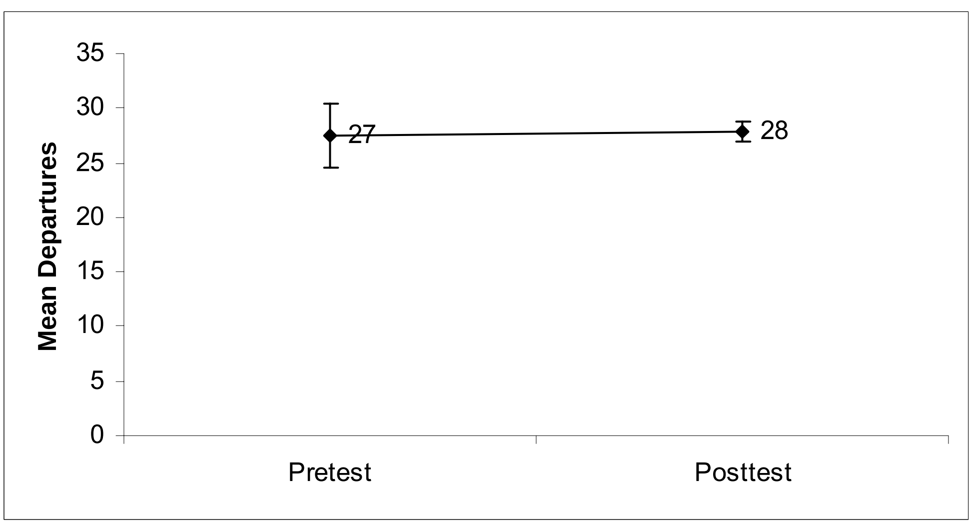


Figure 12. Mean (+/- standard deviation) number of departures as a function of experience.

4.2.2.4 Total Aircraft Handled

We tested whether experience in the simulator had any effect on the total number of aircraft handled in a scenario. We used the total number of aircraft in a scenario for the analysis of the first and last scenarios experienced. The null hypothesis was that the total number of aircraft handled was the same for the pretest and the posttest. The alternative hypothesis was that the total number of aircraft handled was different. We used seven pairs for the permutation test. The test showed that the observed Σd_i (13) was not in the rejection region; therefore, we could not reject the null hypothesis that the total number of aircraft handled was the same. However, inspection of Figure 13 shows that the number of aircraft handled was greater during the posttest. The reader may also note a decline in the variability (pretest *SD* = 2.88, posttest *SD* = .95).

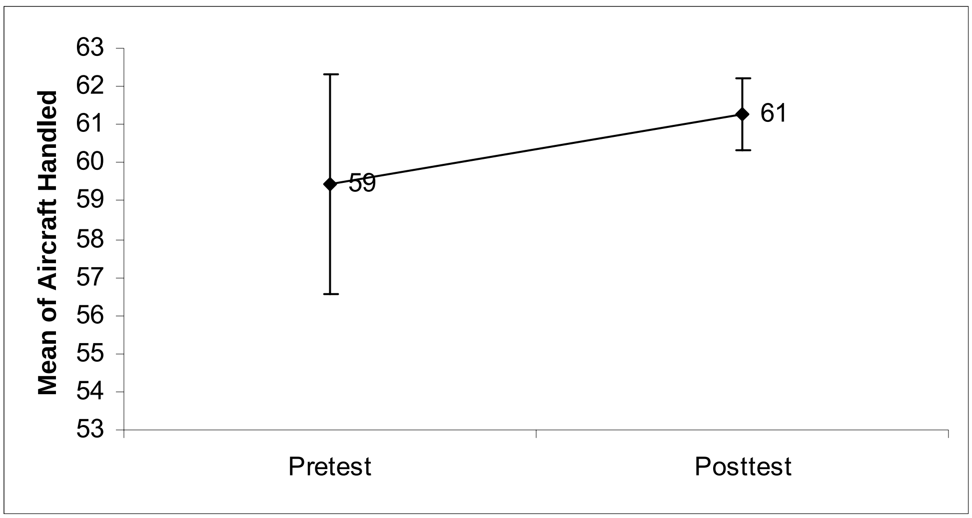


Figure 13. Mean (+/- standard deviation) number of total flights handled as a function of experience.

4.2.3 Safety

Our Safety assessment focused on runway incursions. We included a high number of incursions into the design of the experiment. An incursion occurred if at least one aircraft was approaching within one statute mile at an altitude of less than 500 feet above the runway, or if it was taking off from the runway while one or more taxiing planes were positioned within the runway safety zone. The runway safety zone was 250 feet from the runway centerline. The approach zone was one statute mile from the runway threshold.

By design, controllers were unlikely to have prevented the incursions planned into the scenarios. They only had control over how long the incursions lasted and the closest proximity of the aircraft. We compared the number of incursions that occurred in the first and final scenarios to make certain that the number of incursion were equivalent. Then, we analyzed the duration of the incursions and closest proximity of the aircraft involved. The sample size in each of these analyses was seven, the same that is has been for all analyses.

4.2.3.1 Number of Incursions

The numbers of incursions occurring in the first (pretest) and last (posttest) scenarios were the same. We tested whether the numbers of incursions that occurred in the first and last scenario were different with a paired samples *t* test. The null hypothesis was that the number of incursions was the same; the alternative hypothesis was that the number of incursions in the last scenario was less. The difference did not reach statistical significance, $t(6) = .79$, $p > .05$. The mean number of incursions in the pretest was 11.33 (SD = 2.49) and was 10.14 (SD = 3.02) for the posttest. Therefore, we do not attribute any difference found in the analyses for duration of the incursions and the closest proximity to the mere reduction of incursions in the last scenario.

4.2.3.2 Duration of Incursions

We tested whether experience in the simulator had an effect on the duration of incursions with the permutation test. The null hypothesis tested was that experience in the simulator had no effect on the duration of incursions. The alternative hypothesis was that the durations of incursions in the posttest scenario were less. The observed Σd_i (-18.62) was not in the rejection region and therefore was not statistically significant. The pretest mean, in seconds, was 22.52 (SD = 3.31) and the posttest mean was 25.18 (SD = 11.37).

4.2.3.3 Closest Proximity of Aircraft

We tested whether experience in the simulator had an effect on the closest proximity of aircraft with the permutation test. The null hypothesis was that the closest proximity of the aircraft was the same. The alternative hypothesis was that experience in the simulator resulted in more distance between aircraft in the posttest. The observed Σd_i (6228.24) was not in the rejection region; therefore, we could not reject the null hypothesis. The closest proximity of aircraft in the pretest scenario was 6,742 ft (SD = 968) and in the posttest scenario was 5,252 ft (SD = 2,284).

4.3 Questionnaire Responses

Controllers gave favorable remarks about the simulator as a training tool while also providing a number of complaints about the displays. As shown in Figure 14 to Figure 17 most controllers gave positive ratings to the simulator for use as a training tool, they indicated that it was an improvement over current training methods, and that they would like to use it in the future. However, controllers reported a number of shortcomings with regards to the simulator's realism (See Figure 18 to Figure 20). Specific comments about shortcomings included that the ASDE was unlike the one they used in the tower, the scale of the display was unrealistic, and that it was too cluttered. Regarding the DBRITE, controllers commented: it was unlike the one they used in the tower, it was hard to read, it was in a bad location on the console, it had to be refreshed manually, the range it displayed was inadequate, and that they did not like the double mileage marks on 6R and 6L. In addition, controllers complained that the aircraft were hard to see when they were on final or far away on the ground and that their movement was unrealistic (namely, the movement of the large aircraft was too fast and small aircraft was too slow). Despite these shortcomings, controllers' subjective impression of overall simulation realism (see Figure 21) was generally positive as they indicated they were able to control traffic as they would in the field (Figure 22).

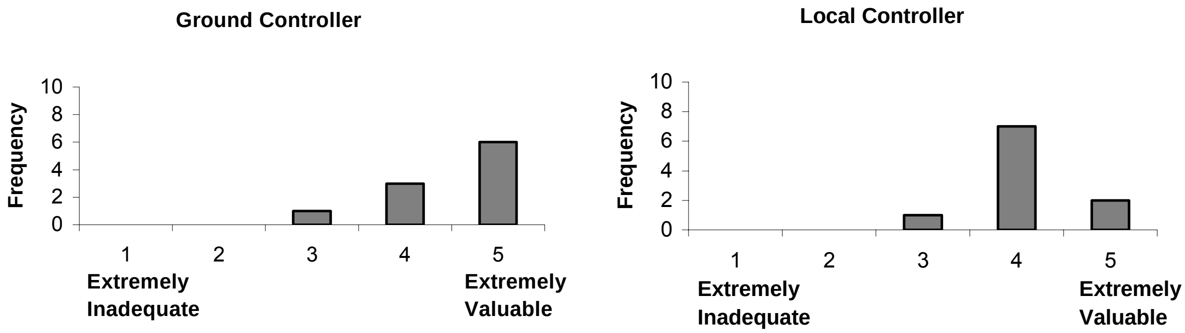


Figure 14. Rate the value of overall simulation experience as a training tool.

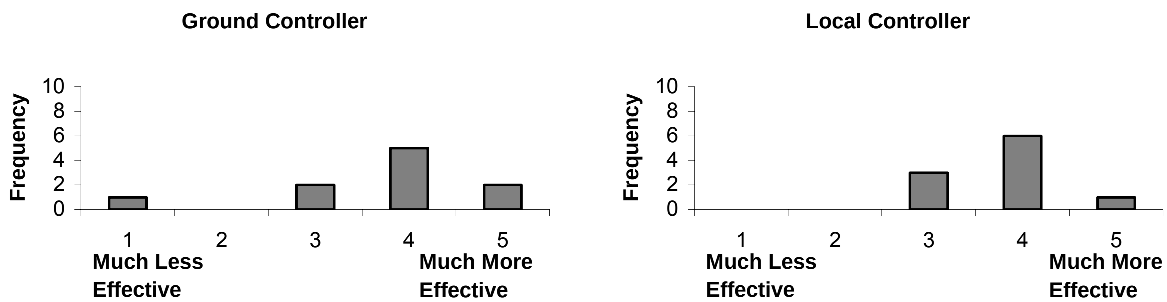


Figure 15. Rate the value of overall simulation experience compared to current tower training methods.

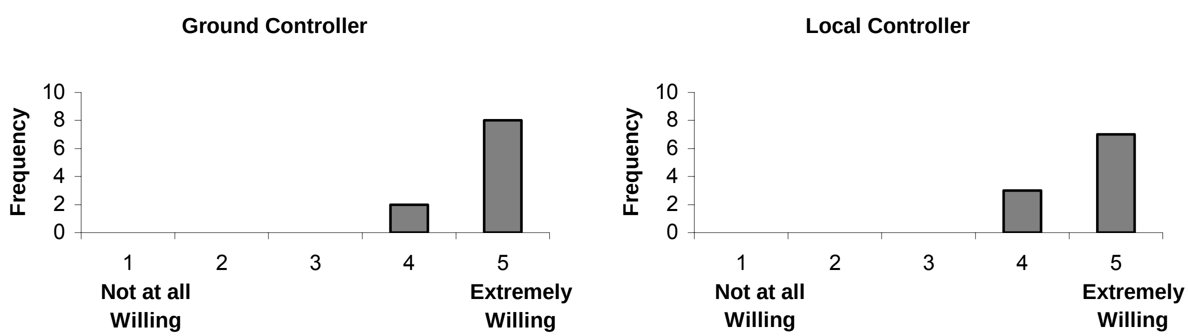


Figure 16. Rate your willingness to participate in high fidelity simulated training in the future.

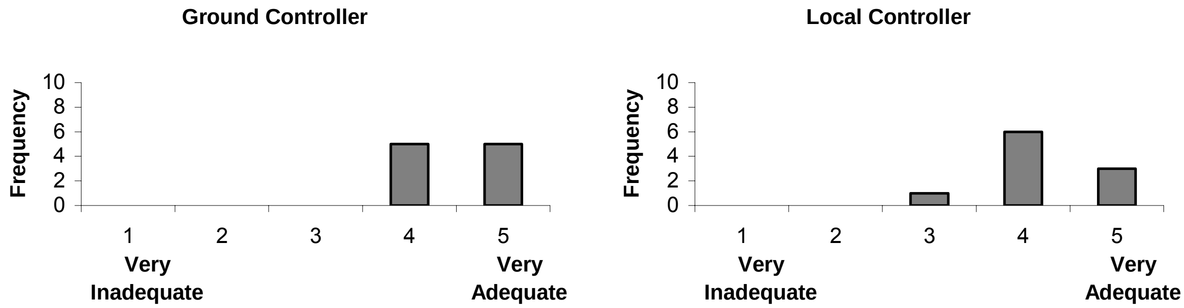


Figure 17. Rate the adequacy of the types of situations (hear-back/read-back errors, losses of separation, etc.) present in the scenarios for training purposes.

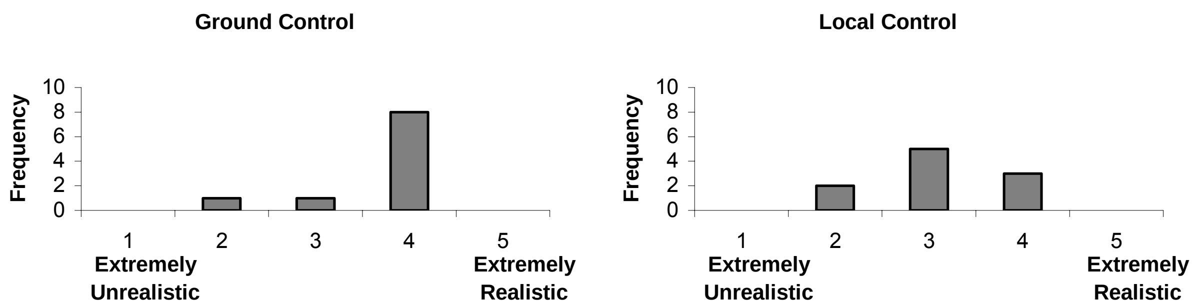


Figure 18. Rate the realism of the overall simulation experience compared to actual control tower operations.

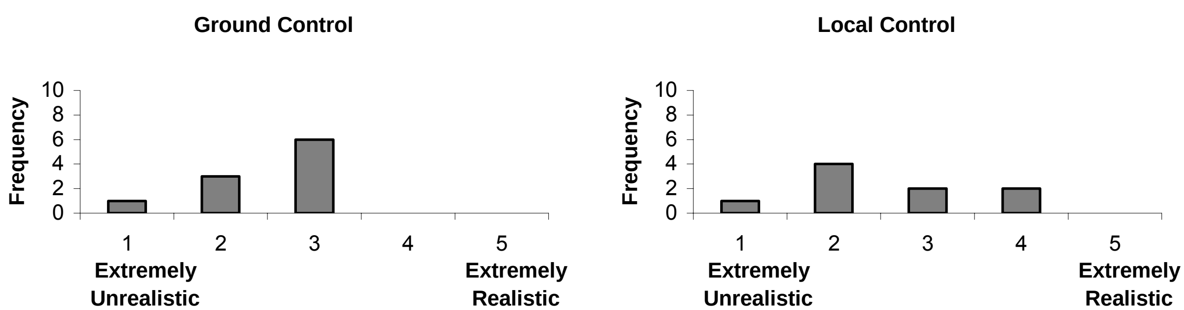


Figure 19. Rate the realism of the simulated hardware (radar displays, communication Equipment, etc.) compared to actual equipment.

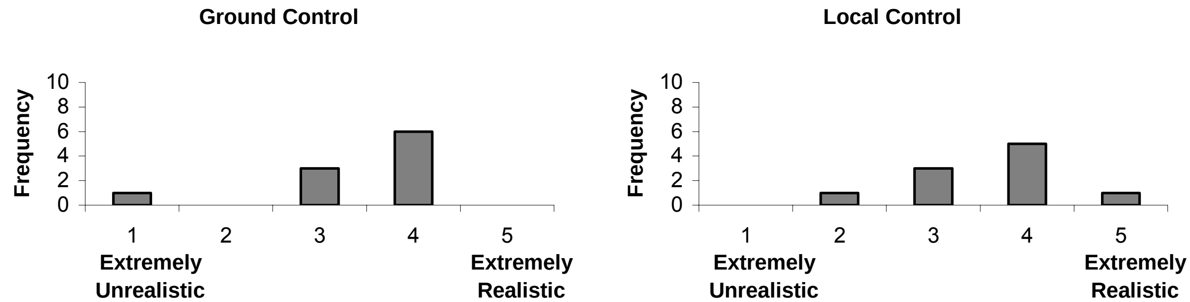


Figure 20. Rate the realism of the simulated software (out the window display, map functionality, etc.) compared to actual functionality.

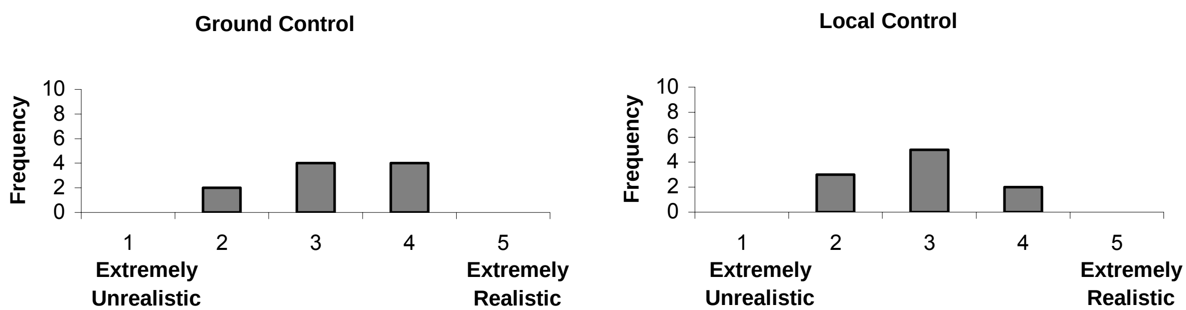


Figure 21. Rate the realism of the overall simulation aircraft behavior compared to actual air traffic operations.

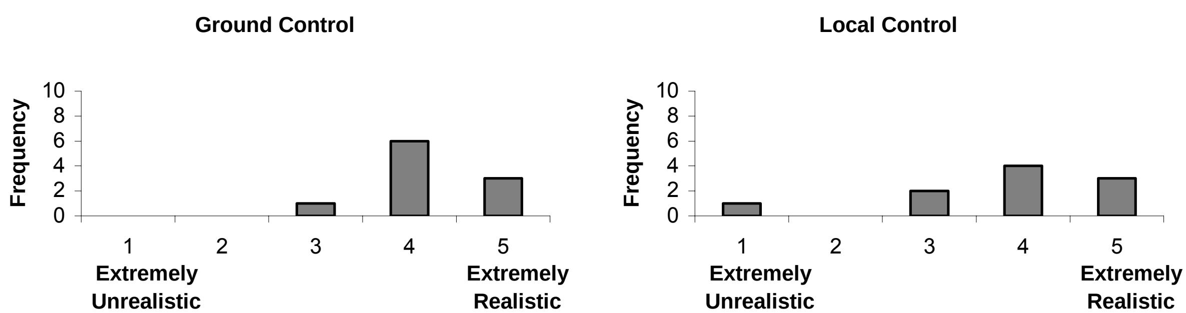


Figure 22. Rate the realism of how you controlled traffic compared to how you would in the field.

5. Discussion

We were able to determine whether exposure to a series of operationally challenging scenarios would result in measurable controller performance improvements. In the following paragraphs, we will discuss our success training hear-back/read-back error detection We will also discuss our failure to train recovery from losses of arrival separation and its potential causes. Finally, we will discuss user acceptance of the simulator as a training tool.

One objective was to train controllers to better recognize hear-back/read-back errors. We were successful doing so using the simulator. The data showed that all but one Ground controller detected every hear-back/read-back error during the posttest.

E/O ratings also indicated that the training for Ground controllers was effective. E/Os indicated that controller responses to hear-back/read-back error-related events in the scenarios improved. In addition, observers saw improvements in separation and control judgment from the first to the last scenario. Improvement in these areas may indicate greater system safety.

The second objective was to train Local controllers to recover from losses of arrival separation. We were unsuccessful training this skill using the simulator. Any combination of the following causes could have contributed to our failure to train controllers successfully recover from losses of arrival separation: improper training strategy, inherent simulator performance limitations, and flawed ATC automation tools.

One limiting source could have been that four scenarios were not adequate to improve performance on this task. Inadequate training may have also been at fault as exposing controllers to losses of arrival separation and providing them with feedback on their handling of them may not have been enough to improve performance. A different approach, such as pausing the simulation after unsuccessful recoveries, may have proved more effective.

It is likely that the limitations of the simulator may have caused the controllers' failure to recover from losses of arrival separation. Typical controller reactions to loss of arrival separation situations are speed reductions or change of runway. Control strategies for using speeds to compensate for aircraft overtakes were often ineffective since the simulator unrealistically limited the speed reductions of large aircraft and speed increases of small aircraft. In addition, the simulator was unable to accept changes to an aircraft's assigned runway within two miles of the threshold. As a result, controller strategies were often ineffective due to these inherent limitations of the simulator, causing an unforeseen reliance on using missed approaches as a solution. However, even the execution of missed approaches was not always successful in maintaining separation. Subsequently, a large number of runway incursions appear in the data.

In addition, there were many complaints made by controllers about the inadequacy of task-related tools (namely, ASDE and DBRITE). In fact, ratings of simulation realism suffered because of these tools. For example, the controllers were unable to increase the range on the DBRITE display; aircraft appeared only 15 miles out on final approach. Controllers were unable to see how the aircraft lined up behind each other, making it difficult to plan ahead. The lack of a runway identifiers on the DBRITE compounded this issue since controllers had to wait for the pilot to report their runway assignment when checking on. Controllers also commented that the double runway range marks (as opposed to range marks on runway 6R only) often contributed to runway assignment confusion. Depiction of the runways on the DBRITE degraded (the approach edge gradually disappeared) if it was not refreshed frequently. This degradation often led controllers to believe there was more time to act, even when aircraft were within two miles of the threshold. This was problematic because of the inability of the simulator to accept changes to an aircraft's assigned runway within this distance.

The simulator gained user acceptance, despite the aforementioned shortcomings. Responses in the questionnaire about the simulator's use as training tool were positive. Controllers were enthusiastic about their participation in this and potential future simulations despite their negative comments about the visibility of aircraft on final and the shortcomings with the console tools. E/Os noted that users were able to overcome these limitations and become absorbed into the problems. They controlled traffic realistically.

6. Conclusion

In this simulation, we were able to show that the tower simulator is useful for training experienced tower controllers. Detection of hear-back/read-back errors improved significantly. System performance measures showed that Efficiency, Capacity, and Safety dimensions remained constant from the first scenario to the last. Although some of the system performance measures did not show statistically significant improvement, slight general trends of improvement (and increases in consistency) were observed in measures such as the time interval between departing aircraft, duration of holds, and throughput. This pattern of results suggests that controllers were able to adapt almost immediately to the simulator environment and focus on the training objectives. Any significant improvements in system performance by these experienced controllers would have been inconsistent with expected expert performance (experts are not expected to show big performance differences) and may have been symptomatic of poor adaptability to the simulator. For example, controllers may have performed less efficiently initially if they had to learn simulator specific procedures not used in the field. In addition, large improvements would not be expected in 45-minutes scenarios. However, small increases in performance or consistency could have a cumulative effect on the efficiency of airport operations. The tower simulator's ability to collect data on system performance and allow real-time E/O and controller participation gives it great potential as a training tool.

7. References

Gawron, V. J. (2000). *Human performance measures handbook*. New Jersey: Lawrence Earlbaum Associates.

Federal Aviation Administration. (2002) *Runway incursion joint safety implementation team: Results and analysis.* Washington, DC

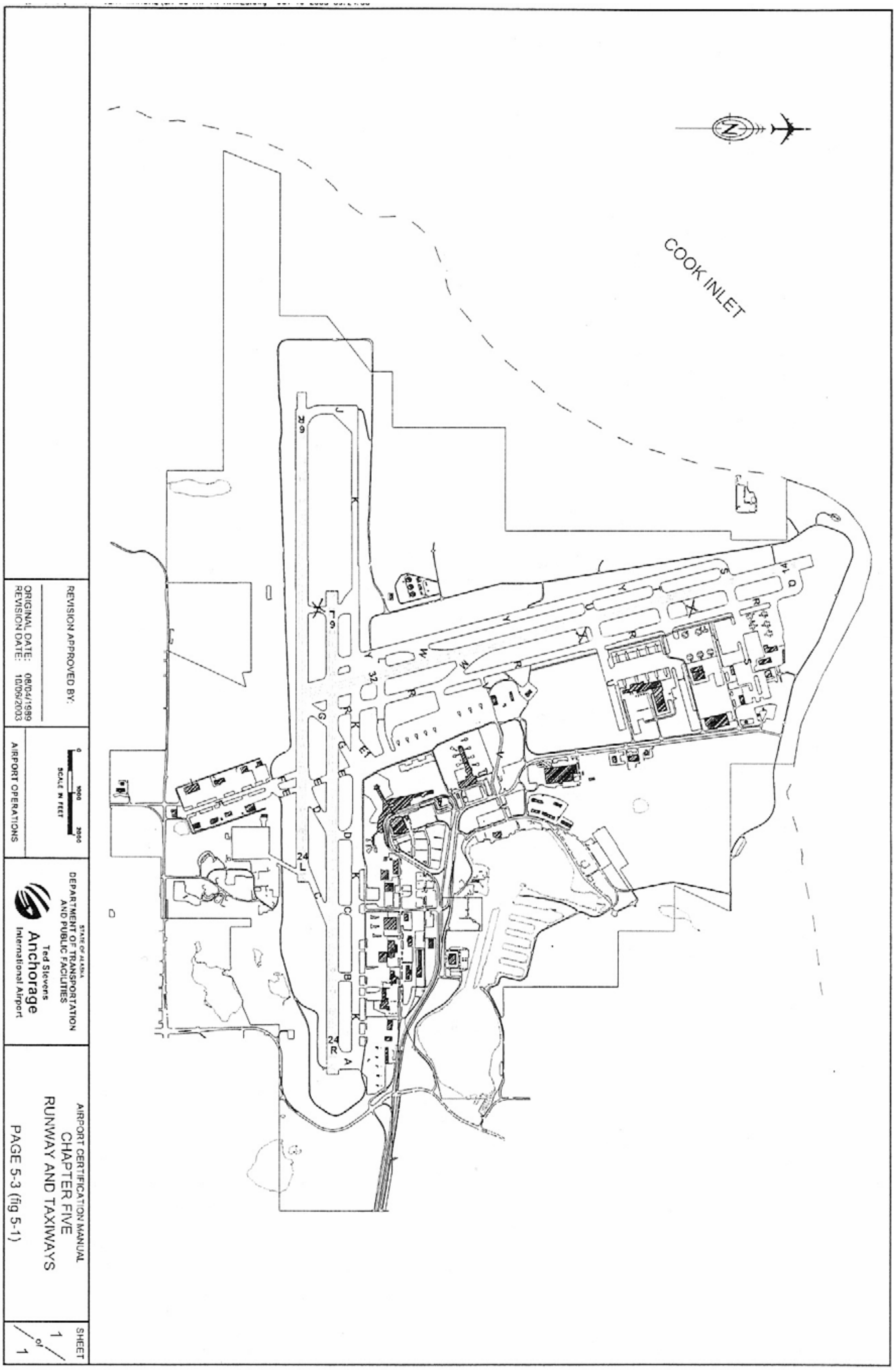
Federal Aviation Administration. (2003) *Federal Aviation Administration Flight Plan 2004-2008.* Washington, DC. Retrieved March 4, 2004, from http://www.faa.gov/AboutFAA/FlightPlan.cfm

Hart, S. G., and 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). North Holland, Amsterdam, Netherlands

Siegel, S., & Castellan, N. J., Jr. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.

APPENDIX A – TSAIA AIRPORT DIAGRAM

TSAIA Airport



APPENDIX B – SAMPLE E/O OVER-THE-SHOULDER FORMS

Scenario 5 **Controller ID: ______** **Date: _________** **Observer ID: _______**

Position: Local

Task	Subtask	1-Never	2-Seldom	3-Sometimes	4-Often	5-Always
Separation	1. Separation is ensured					
Coordination	2. Required coordination is performed.					
Control Judgment	3. Good control judgment is applied.					
	4. Priority of duties is understood.					
	5. Positive control is provided.					
	6. Effective traffic flow is maintained.					
Methods and Procedures	7. Aircraft identity is maintained.					
	8. Strip posting is complete/correct.					
	9. Clearance delivery is complete/correct and timely.					
	10. LOAs/directives are adhered to.					
	11. Additional services are provided.					
	12. Rapidly recovers from equipment failures and emergencies.					
	13. Scans entire control environment.					
	14. Effective working speed is maintained.					
Equipment	15. Equipment capabilities are utilized/understood.					
Communication	16. Functions effectively as a radar/tower team member.					
	17. Communication is clear and concise.					
	18. Uses prescribed phraseology.					
	19. Makes only necessary transmissions.					

COMMENTS

____

____

____

____

____

____

____

____

____

CRITICAL ACTION POINT 5-L__-1 **TOC**

DESCRIPTION N973M (C550) land runway 6R (12 minutes into the problem) programmed to exit at Foxtrot (SAP) intersection and UPS6927 (H/B747) landing 6R.

---

1 PLEASE CHECK ONE

- □ Completely unobserved
- □ Observed but no action taken
- □ Action taken at 1st opportunity…………….. □ succeeded □ failed
- □ Action taken at 2nd opportunity point……… □ succeeded □ failed

---

		PLEASE CIRCLE ONE				
2	RECOGNITION TIME The amount of time taken to recognize the event	1 Very inadequate	2 Slightly inadequate	3 Borderline	4 Slightly adequate	5 Very adequate
3	APPROPRIATENESS OF ACTION The appropriateness of the actions taken to resolve the event	1 Very inappropriate	2 Slightly inappropriate	3 Borderline	4 Slightly appropriate	5 Very appropriate
4	APPROPRIATENESS OF TIMING The appropriateness of the timing of the actions taken to resolve the event	1 Very inappropriate	2 Slightly inappropriate	3 Borderline	4 Slightly appropriate	5 Very appropriate
5	ACCURACY OF COMMUNICATION - TRAFFIC The accuracy of communication with traffic regarding the event	1 Very unacceptable	2 Slightly unacceptable	3 Borderline	4 Slightly acceptable	5 Very acceptable
6	ACCURACY OF COMMUNICATION - ATC The accuracy of communication with other controllers regarding the event	1 Very unacceptable	2 Slightly unacceptable	3 Borderline	4 Slightly acceptable	5 Very acceptable
7	OVERALL RATING The overall adequacy of the response to the event	1 Very inadequate	2 Slightly inadequate	3 Borderline	4 Slightly adequate	5 Very adequate

*potential event and actual event used interchangeably

8 COMMENTS

____

____

____

____

APPENDIX C – PERFORMANCE DATA

Performance Data Reported

MEASURE	DESCRIPTION
Expert Observation	
Over-the-Shoulder Form	E/O ratings of controller performance on several dimensions including separation, coordination, control judgement, methods and procedures, equipment, and communications
CAP Ratings Form	E/O accounts of controller responses to planned events in a scenario including recognition of the event and performance ratings
Efficiency	
Time interval between landing aircraft	Time interval between landing aircraft for individual runways
Time interval between departing aircraft	Time interval between departing aircraft for individual runways
Number of missed approaches	Frequency count of aircraft that failed to land
Taxi time	Duration of taxi in and taxi out
Duration of holds	Time that aircraft held on the runway before takeoff
Capacity	
Throughput	Count of aircraft from pushback to takeoff and landing to gate
Arrivals	Number of aircraft that touched down in a scenario
Departures	Number of aircraft that took off in a scenario
Total aircraft handled	Total number of aircraft introduced into the scenario
Safety	
Number of incursions	Occurred if at least one aircraft is approaching within one statute mile at an altitude of less than 500 feet above the runway or is taking off from the runway while one or more taxiing planes were positioned within the runway safety zone. The runway safety zone is 250 feet from the runway centerline. The approach zone is one statute mile from the runway threshold.
Duration of incursions	Time that aicraft were in a position defined as an incursion
Closest proximity of aircraft	Least distance between aircraft involved in an incursion

APPENDIX D – SAMPLE SCENARIO CRITICAL ACTION POINT DESCRIPTION

Detailed Scenario Events

CRITICAL ACTION POINTS

Position-Type Error	(Number) and Type of Critical Action Point
LC	
LC-1	**(4-5) Loss of arrival separation with successive arrivals.** Les has identified and Shaun is to program in spacing between certain arrivals so that, when the first arrival turns to exit at the predetermined taxiway, there will not be adequate runway separation between it and the second arrival.
LC-2	**(2-4) Hear-back/Read-back errors on "Hold short of a runway" clearance**. This will be either aircraft told to "hold short" of the parallel runway after landing, or told to hold short of their departure runway. These will not necessarily result in a runway incursion.
LC-3	**(1) Pilots crossing a runway or taxiing on a runway after acknowledging "hold short" instructions.** These will result in a runway incursion and if caught, just about the only option is a "go around".
LC-4	**(1) Pilots deviating from normal performance.** (i.e. - Pilots taking longer than expected for take-off roll, or crossing a runway.)
N/A	(undefined) **Teamwork** Help catching GC's Critical Action Points.

GC	
GC-1	**(3-5) Hear-back/Read-back errors on "Hold short of a runway" clearance**. These will (mostly) be aircraft that are taxiing for runway 32 extension, told to hold short of runway 6 Left, but do not read-back the "hold short" instructions. These will not necessarily result in a runway incursion.
GC-2	**(1) Pilots getting lost.** These will not necessarily result in a runway incursion.
	(undefined) **Teamwork** – Help catching LC's Critical Action Points.

Scenario ANC ATCT 01

Local Control

Loss of arrival separation with successive arrivals (4-5)

PEN63 (SF34) land runway 6R (14 minutes into the problem) programmed to exit at Delta intersection and FDX96 (H/DC10) landing 6R.

Scars55 (C130) land runway 6R (24 minutes into the problem) programmed to exit at Charlie intersection (Kulis) and ASA85 (B737) landing 6R.

ASA93 (B737) land runway 6R (28 minutes into the problem) programmed to exit at Echo intersection and JAL6401 (H/B747) landing 6R.

GTI531 (H/B747) land runway 6R (37 minutes into the problem) programmed to exit at Echo intersection and ERH807 (DH8) landing 6R.

PEN2081 (SW4) land runway 6L (48 minutes into the problem) programmed to exit at Delta intersection and ERH897 (DH8) landing 6L.

Hear-back/Read-back errors on “Hold short of a runway” clearance (2-4)

ERH880 (DH6) departs runway 6L (11 minutes into the problem) If the controller issues a “hold short of runway 6L” clearance, , the aircraft will not read back the “hold short instructions.

PAC9000 land runway 6R (45 minutes into the problem) If the controller issues a “hold short of runway 6L” clearance, , the aircraft will not read back the “hold short instructions.

Pilots crossing a runway after acknowledging “hold short” instructions (1)

UPS2998 (H/B747) lands runway 6R (42 minutes into the problem) will read back the “hold short instructions correctly, but will immediately cross runway 6L in front of landing and/or departing traffic.

Pilots deviating from normal performance (1)

PAC148 (H/B747) departs runway 32 from the extension (39 minutes into the problem) PAC 148 will (unless otherwise prompted by the controller) sit for 2 minutes on the extension before starting takeoff roll.

Scenario ANC ATCT 01

Ground Control

Hear-back/Read-back errors on “Hold short of a runway” clearance (3-5)

CAL322 taxiing for runway 32 from the extension (24 minutes into the problem) will not read back the “hold short instructions.

PAC148 taxiing for runway 32 from the extension (39 minutes into the problem) will not read back the “hold short instructions.

CSN491 taxiing for runway 32 from the extension (50 minutes into the problem) will not read back the “hold short instructions.

Pilots getting lost (1)

N71911 (C206) taxiing from Lake Hood to runway 32 will get lost and taxi west on Victor Taxiway, then north on Romeo Taxiway to runway 14 at Tango intersection. The aircraft will call the Tower ready for departure.

APPENDIX E – CONSENT FORM

Participation Consent Form

To the Research Participant: Please read this consent form and the attached protocol and/or subject instructions carefully. Make sure all of your questions have been answered to your satisfaction before signing.

I agree to participate in the Advanced Controller Training in a Virtual Environment-1 study. I understand that the Federal Aviation Administration (FAA) Office of Runway Safety sponsors this study, Alaskan Region Runway Safety Office is the project manager, and that the FAA's Simulation and Analysis Group (ACB-330) directs this study.

Nature and Purpose:
The ACTIVE-1 study is a training feasibility study that is intended to provide an initial examination of the suitability of high fidelity, state-of-the-art tower simulators to provide recurrent training for current and qualified controllers. Controllers will be exposed to a series of operationally challenging scenarios in the tower simulator and researchers will look for measurable performance improvement. This study will only examine performance as it relates to the training method and is not intended to evaluate individual controller performance.

Experimental Procedures:
Twelve Certified Professional Controllers (CPCs) with TSAIA experience will participate in the simulation over a two-week period. Controllers are scheduled to participate one to two days. Each controller will experience five simulation runs and will operate the Local or Ground position. Controllers that participate for two days will operate both positions. Subjective and objective measures will be collected during each simulation run.

Discomforts and Risks:
There are no expected discomforts or risks associated with this simulation.

Benefits:
I understand that the benefit to me is the opportunity to participate in research that examines the feasibility of tower simulation training on all controllers.

Participant Responsibilities:
During the simulation it will be my responsibility to control air traffic and regard the simulated air traffic as if it were live traffic. I will answer any questions asked during the simulation to the best of my abilities. I will not discuss the content of the simulation with anyone until its formal completion. I will complete a background questionnaire, a post-run questionnaire at the end of each simulation run, and a post-simulation questionnaire at the end of all simulation runs. I will participate in debriefs at the end of each simulation run, and at the completion of the full simulation.

Participant's Assurances:
I understand that my participation in this simulation is completely voluntary. The Principal Investigator will adequately answer any and all questions I have about this simulation, my participation, and the procedures involved. I understand that if new findings develop during the course of this research that may relate to my decision to participate, I will be informed.

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

I understand that records of this simulation are strictly confidential, and that I will not be identifiable by name or description in any reports or publications about this simulation. Video and audio recordings are for use within the William J. Hughes FAA Technical Center (WJHTC) only. Any of the materials that may identify me as a participant cannot be used for purposes other than internal to the WJHTC without my written permission.

I understand that I can withdraw from the simulation at any time without penalty or loss of benefits to which I may be entitled. I also understand that the researcher or sponsor of this simulation may terminate my participation if he or she feels this to be in my best interest.

If I have questions about this simulation or need to report any adverse effects from the research procedures I will contact Nicole Racine at (609) 625-5669.

I have read this participation form, I understand its contents, and I freely consent to participate in this simulation under the conditions described. I have received a copy of this participation form.

Signature of Research Participant: ____________________Date: ______________

Research Director: _________________________________Date: ______________

Witness: ______Date: _

APPENDIX F - BACKGROUND QUESTIONNAIRE

BACKGROUND QUESTIONNAIRE

The following information is requested for reporting data relevant to the ACTIVE-1 simulation.

Your personal information will be kept completely confidential and will not be included in any of the reports or documents that will be produced as a result of this study. When necessary, individuals will be identified as Subject L_1, Subject L_2, etc.

Participant Code: __________

Date: ____________________

---

1. Are you a Certified Professional Controller?

Circle one: YES NO

2. What is your total experience as a CPC controller (in any control position and geographic location)?

Years: ______________ Months: ______________

3. What is your total experience as a CPC Tower controller (in any control position and geographic location)?

Years: ______________ Months: ______________

4. What is your total experience as a CPC TSAIA Tower controller?

Years: ______________ Months: ______________

APPENDIX G – POST RUN QUESTIONNAIRE

Scenario ID ________ **Date:** _________
Controller ID: ________
Position: Ground or Local

Compare Scenarios

	PLEASE CIRCLE ONE		
1. SCENARIO WORKLOAD: The workload demands of this scenario compared to the last scenario	Much Less	① ② ③ ④ ⑤	Much More
2. SCENARIO COMPLEXITY: The complexity of this scenario compared to the last scenario	Much Less	① ② ③ ④ ⑤	Much More
3. SCENARIO DIFFICULTY: The difficulty of this scenario compared to the last scenario	Much Less	① ② ③ ④ ⑤	Much More
4. FEEDBACK: The feedback given after this scenario compared to the last scenario	Much Worse	① ② ③ ④ ⑤	Much Better

Training

Instructions: For this section "training" refers to experiencing the scenarios in today's simulation.

5. TRAINING-Situation Detection: The adequacy of the amount of practice recognizing factors leading to events	Very Inadequate	① ② ③ ④ ⑤	Very Adequate

6. If your rating for question 5 was "3" or below, please identify which error types would you benefit from more practice? (Loss of arrival/departure separation; miscommunications-hear-back errors; mistaken aircraft/vehicle locations or cleared to land, other

____
____
____
____

7. Which have you already mastered?

____
____
____
____

8. TRAINING-Situation Response: The adequacy of the amount of practice responding/handling events

Very Inadequate ① ② ③ ④ ⑤ Very Adequate

9. If you gave a rating of 3 or below for question 7, with which events would you benefit from more practice?

___

___

___

___

10. Which have you already mastered?

___

___

___

___

APPENDIX H – POST SIMULATION QUESTIONNAIRE

POST SIMULATION QUESTIONNAIRE

Participant Codes:________________	**Date: _____/_____/ 2004**

Instructions: All questions are specific to this simulation. Please answer the following based upon your overall experience in the simulation. As always, your identity will remain anonymous.

Familiarization

Instructions: Familiarization refers to the initial run providing the opportunity to acclimate to the simulator environment.

1. Rate the **adequacy of the familiarization runs** you received for the simulation.	Extremely Poor	① ② ③ ④ ⑤	Extremely Good

2. If your rating on question 1 was a "3" or below, please identify the familiarization deficiencies and describe what should be done to improve (e.g., longer practice sessions, more robust scenarios, etc.).

____

____

____

Overall Simulator Training

Instructions: Overall Simulator Training refers to the entire simulation experience including familiarization, instruction, scenario runs, questionnaires, feedback, and debriefing.

3. Rate the **value of overall simulation experience** as a training tool.	Extremely Inadequate	① ② ③ ④ ⑤	Extremely Valuable

4. If your rating on question 3 was a "3" or below, please describe what you think could improve the simulator as a training tool.

____

____

____

____

5. Rate the **value of overall simulation experience** compared to current tower training methods.	Much Less Effective	① ② ③ ④ ⑤	Much More Effective

6. Rate the **effectiveness of the scenarios** as a training tool.	Extremely Ineffective	① ② ③ ④ ⑤	Extremely Effective

7. If your rating on question 6 was a "3" or below, please describe the type of training (instead of the scenarios you experienced) that you would consider effective?

____

____

8. Rate the **degree to which you could apply the simulation experience** to actual control tower operations.	Extremely Inapplicable	① ② ③ ④ ⑤	Extremely Applicable

9. Rate your **level of alertness** to possible unusual events compared to actual control tower operations	Greatly Decreased	① ② ③ ④ ⑤	Greatly Increased

10. Rate your **willingness to participate** in a hi-fidelity simulated training in the future.	Not at all Willing	① ② ③ ④ ⑤	Extremely Willing

Feedback Effectiveness

Instructions: Feedback refers to the post-run questionnaires, debriefing, and discussion of scenario events.

11. Rate the adequacy of the **amount of feedback** provided after each scenario in terms of how it helped you meet the training objectives. For example, did you find the amount of time debriefing, reviewing the events, and the amount of time discussing cues leading to those events adequate.	Very Inadequate	① ② ③ ④ ⑤	Very Adequate

12. If your rating on question 11 was a "3" or below, please describe the amount and/or type of feedback you would find adequate.

____
____
____

Training Effectiveness

Instructions: For this section "training" refers to experiencing the scenarios in today's simulation.

13. Rate the adequacy of the **type of training** you received. For example, did you find that having different types of events within each scenario gave you the type of training experience you needed?	Very Inadequate	① ② ③ ④ ⑤	Very Adequate

14. If your rating on question 13 was a "3" or below, please describe the type of practice you would find adequate.

____
____
____

15. Rate the adequacy of the **amount of training** you received. For example, did you find that the number of scenarios gave you enough training experience?	Very Inadequate	① ② ③ ④ ⑤	Very Adequate

16. If your rating on question 15 was a "3" or below, please describe the amount of practice you would find adequate.

____
____

17. Rate your **level of alertness** to possible unusual events compared to actual control tower operations	Greatly Decreased	① ② ③ ④ ⑤	Greatly Increased

18. Do you have any comments or suggestions for improvement to the training in the tower simulator?

____
____
____
____
____

Overall Simulation Fidelity

19. Rate the **realism of the overall simulation experience** compared to actual control tower operations.	Extremely Unrealistic	① ② ③ ④ ⑤	Extremely Realistic
20. Rate the **realism of the simulated hardware (radar displays, communication equipment, etc.)** compared to actual equipment.	Extremely Unrealistic	① ② ③ ④ ⑤	Extremely Realistic
21. Rate the **realism of the simulated software (out the window display, map functionality, etc.)** compared to actual functionality.	Extremely Unrealistic	① ② ③ ④ ⑤	Extremely Realistic
22. Rate the **realism of the overall simulated controller communications** compared to actual control tower operations.	Extremely Unrealistic	① ② ③ ④ ⑤	Extremely Realistic
23. Rate the **realism of the overall simulated pilot communications** compared to actual control tower operations.	Extremely Unrealistic	① ② ③ ④ ⑤	Extremely Realistic
24. Rate the **realism of the overall simulation aircraft behavior** compared to actual air traffic operations.	Extremely Unrealistic	① ② ③ ④ ⑤	Extremely Realistic
25. Rate the **realism of the simulated traffic runs** compared to actual National Airspace System (NAS) traffic.	Extremely Unrealistic	① ② ③ ④ ⑤	Extremely Realistic
26. Rate the **realism of the simulated airport** compared to the actual airport.	Extremely Unrealistic	① ② ③ ④ ⑤	Extremely Realistic
27. Rate the **realism of how you controlled traffic** compared to how you would in the field.	Extremely Unrealistic	① ② ③ ④ ⑤	Extremely Realistic

28. Do you have any comments or suggestions for improvement about the simulation capability?

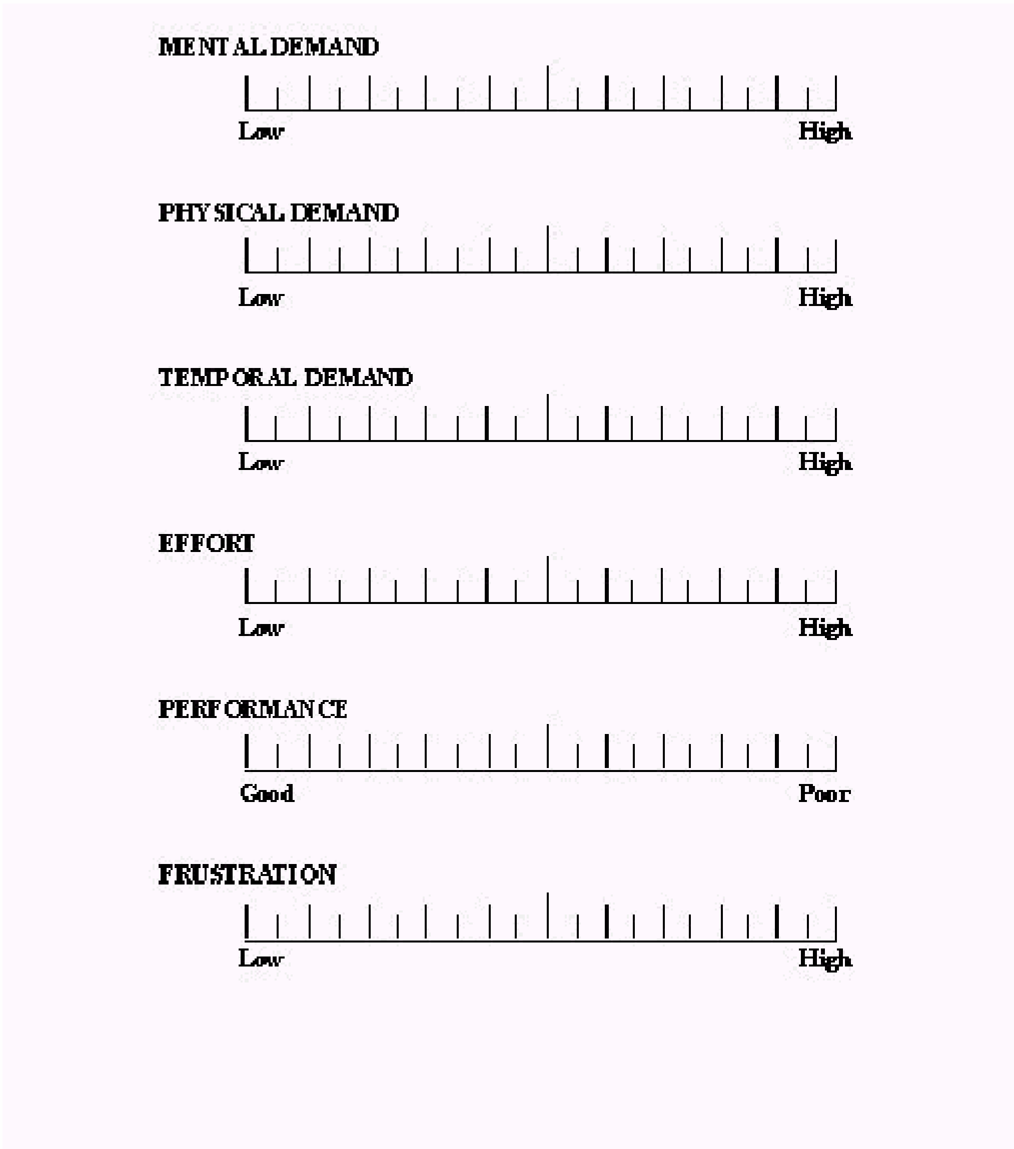
____
____
____
____
____

Comments

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

APPENDIX I – NASA TLX

Scenario ______ **Controller ID:____________** **Date:_________**



RATING SCALE DEFINITIONS		
Title	Endpoints	Descriptions
MENTAL DEMAND	Low/High	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	Low/High	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	Low/High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
EFFORT	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
PERFORMANCE	Good/Poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
FRUSTRATION LEVEL	Low/High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?