technical note technical note

The Development and Evaluation of a Behaviorally Based Rating Form for the Assessment of En Route Air Traffic Controller Performance

Jennifer J. Vardaman, Ph.D., PERI
Earl S. Stein, Ph.D., ACT-530

June 1998

DOT/FAA/CT-TN98/5

U.S. Department of Transportation
Federal Aviation Administration

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

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents or use thereof.

The United States Government does not endorse products or manufacturers. Trade or manufacturers' Names appear herein solely because they are considered essential to the objective of this report.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/CT-TN98/5		

4. Title and Subtitle	5. Report Date
The Development and Evaluation of a Behaviorally Based Rating Form for the Assessment of En Route Air Traffic Controller Performance	June 1998
	6. Performing Organization Code ACT-530
7. Author(s) Jennifer J. Vardaman, Ph.D., PERI and Earl S. Stein, Ph.D., ACT-530	**8. Performing Organization Report No.** DOT/FAA/CT-TN98/5
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. F2202K
12. Sponsoring Agency Name and Address Federal Aviation Administration Human Factors Division 800 Independence Ave., S.W. Washington, DC 20591	**13. Type of Report and Period Covered** Technical Note
	14. Sponsoring Agency Code AAR-100

15. Supplementary Notes

16. Abstract

This project expanded and evaluated the performance evaluation method developed by Sollenberger, Stein, and Gromelski (1997), a Terminal Radar Approach Control rating form and training package designed to better assess air traffic controller performance. The form is a research-oriented testing and assessment tool designed to measure the efficacy of new air traffic control (ATC) systems, system enhancements, and operational procedures in simulation research. The rating form used in the present study focused on observable behaviors that supervisory air traffic control specialists (SATCSs) use to make behaviorally based ratings of en route controller performance. The present study evaluated the inter-rater and intra-rater reliability of performance ratings made by nine Air Route Traffic Control Center supervisors who viewed videotapes and computerized replays of controllers from a previously recorded en route study. The rating form contained 26 items, which were organized into six major categories. Various observable behaviors, which SATCSs identified as those they consider when assessing controller performance, anchored each performance area. Inter-rater (between rater) reliability of SATCS performance ratings assessed using intra-class correlations was somewhat low. Intra-rater (within rater) reliability of SATCS performance ratings was consistent with previous studies and indicated that raters were stable over time in the ratings they assigned. Researchers also investigated the relationship between SATCS performance ratings and personality traits from the Sixteen Personality Factor personality inventory. The results indicated that what SATCSs bring with them to the experimental evaluation setting, in terms of personality traits, may be related to their ratings. Future research efforts should concentrate on distinguishing the sources of measurement error and making whatever changes necessary to produce a reliable controller performance assessment tool.

17. Key Words	18. Distribution Statement
En Route Air Traffic Control Controller Performance Assessment Air Traffic Control Simulation	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	65	

Acknowledgments

The authors wish to acknowledge several people who contributed to this study. Dave Cognata, Supervisory Air Traffic Control Specialist, Jacksonville Air Route Traffic Control Center, served as the subject matter expert on the project. Dr. Laurie Davidson, Princeton Economic Research, Inc., (PERI), audio recorded the training sessions and provided significant feedback on the comments made by the participants during the study. Albert Macias and Mary Delemarre, ACT-510, tirelessly worked to ensure that ATCoach ran to the satisfaction of the researchers. George Rowand, System Resources Corporation, spent countless hours editing videotapes and synchronizing the audio and videotapes each day of the study. Bill Hickman, ACT-510, and Mike Cullum and Bruce Slack, System Resources Corporation, served as the simulation pilots on the project. Dr. Earl Stein, ACT-530, served as the project lead for this study. Dr. Randy Sollenberger, ACT-530, assisted with participant training and drafting the test plan. Finally, Paul Stringer, Vice President - Aviation Division, PERI, also contributed to the design of this study.

Table of Contents

Table of Contents (Cont.)

Appendixes

List of Illustrations

Tables

Executive Summary

In this second study on performance rating, researchers investigated the process used by supervisory air traffic control specialists (SATCSs) to rate en route air traffic control specialists (ATCSs). This project expanded and evaluated an earlier performance evaluation method developed for Terminal Radar Approach Control (TRACON) ATCSs. This rating form and training package was a testing and assessment tool to measure the efficacy of new air traffic control systems, system enhancements, and operational procedures in simulation research.

The rating form used in the present study focused on observable en route behaviors that SATCSs can use to make behaviorally based ratings of controller performance. The present study evaluated the reliability of the rating process by determining the level of agreement between ratings of air route traffic control center (ARTCC) supervisors who viewed videotapes and computerized graphical replays of controllers from a previously recorded en route study.

The en route rating form contained 26 items. However, participants concluded that they had insufficient information to rate two items. The performance areas were organized into six categories: Maintaining Safe and Efficient Traffic Flow, Maintaining Attention and Situational Awareness, Prioritizing, Providing Control Information, Technical Knowledge, and Communicating. Observable behaviors anchored each performance area. SATCSs identified these behaviors as those they consider when assessing ATCS performance. The rating form contained an eight-point rating scale format with statements describing the applicable controller actions for each point. A comment section for each item provided space for participants to explain the ratings they assigned.

The study took place in the Research Development and Human Factors Laboratory (RDHFL) at the Federal Aviation Administration William J. Hughes Technical Center, Atlantic City International Airport, New Jersey. Nine en route SATCSs from five different ARTCCs participated as observers. The RDHFL video projection system presented three views of a previously recorded en route study. The primary view was a graphical playback of the traffic scenario that showed all the information on the controller's radar display. Another view was an over-the-shoulder video recording of the controller's upper body that showed interactions with the workstation equipment. The third view was a video recording of the traffic scenario as it appeared on the simulation pilot's display. All three views were simultaneously presented on different screens and synchronized with an audio recording of the controllers and simulation pilots.

The researchers assessed two types of reliability: inter-rater and intra-rater. Inter-rater reliability refers to the uniformity of the ratings between participants, and intra-rater reliability refers to the uniformity of the ratings on repeated occasions.

The results of the present study indicated that the inter-rater reliability of the en route rating form ranged from $r = .27$ to $r = .74$. The overall ratings for each performance category were generally more reliable than the individual ratings included in each category. The intra-rater reliabilities were higher. Participants were more consistent individually over time than they were between each other reviewing the same controller behavior.

There are possible explanations for the inter-rater reliability coefficients. Participants concluded that they had specialized knowledge and wanted to take a very active role in the process of developing the rating form and its associated training. Second, the changes, even though recommended by the en route SATCSs who participated in the present study, may also have had an impact on inter-rater reliability. Finally, there were some problems with the simulation replay technology during the present study.

Researchers also investigated the relationship between participant ratings and selected personality traits. Participants completed the Sixteen Personality Factor (16PF) personality inventory. The results indicated that the personality traits participants bring with them to the experimental setting may be related to their ratings. Such traits are difficult to overcome with only 1 week of training in the experimental environment.

The performance rating form is a research-oriented assessment tool, which provides data about controller performance that is not available from other sources. Future research efforts should focus on identifying the sources of measurement error and making whatever changes are necessary to produce a more reliable instrument.

1. Introduction

This is the second in a series of research studies involved in developing more effective performance rating procedures. The first study involved developing a performance rating form to test and assess simulation research using Terminal Radar Approach Control (TRACON) personnel. The present study concentrated on a performance rating form for en route air traffic control specialists (ATCSs).

1.1 Background

Sollenberger, Stein, and Gromelski (1997) conducted the first study. They developed the TRACON rating form to assess new air traffic control (ATC) systems, system enhancements, and operational procedures. They attempted to (a) build a reliable tool for measuring controller performance in the research setting; (b) improve the quality of ATC performance evaluations; (c) improve the quality, reliability, and comprehensiveness of ATC evaluations and tests in the research setting; and (d) identify criteria for evaluating controller performance.

The Sollenberger et al. (1997) study indicated that the rating process was workable in a TRACON environment. It also identified the performance areas that were more difficult for participants to evaluate consistently, possibly due to misunderstanding rating criteria or overlooking critical controller actions. Finally, the study demonstrated the feasibility of using video and computerized graphical playback technology as a presentation method for evaluating controller performance.

1.2 Problem Statement

Human performance is essential to overall system performance. The decisions humans make and how they act on them directly impact the degree to which the system achieves its goals. There is, however, disagreement on what role the human plays in the system and what makes up human performance. Most systems have some definition of minimum essential performance for their human operators, but they do not distinguish levels of performance quality above the minimum level. The problem, then, is, if standards of performance are not well defined, how do subject matter experts (SMEs) know what constitutes “acceptable” or “unacceptable” performance?

Researchers at the Federal Aviation Administration (FAA) William J. Hughes Technical Center have studied human performance issues for many years. Much of this research has stressed system effectiveness measures (SEMs) that can be collected in real time during ATC simulations. SEMs are objective measures that can be collected and analyzed to assess the effects of new ATC systems and procedures on controller performance.

1.3 Assumptions and Goals

Sollenberger et al. (1997) conducted a study to determine if SEMs are related to how SMEs evaluate controller performance. The authors investigated whether or not SMEs could be trained to evaluate ATCS performance so that they were looking at the same behaviors and assigning similar values to them. They also investigated whether or not SMEs’ combined performance evaluations are related to the SEMs, assuming that the SMEs ratings are reliable.

Sollenberger et al. (1997) believed that it is possible to train supervisory ATCSs (SATCSs) to objectively observe and evaluate controller performance. SATCSs are experienced with FAA Form 3120-25, the ATCT/ARTCC OJT Instructor Evaluation Report. The authors assumed that FAA Form 3120-25 could be improved, and, when supported by a training curriculum, performance-rating quality would also improve. They did not intend to develop a performance evaluation form to replace FAA Form 3120-25. Rather, they intended to develop an observational performance rating system that could be used to validate other measurement systems.

Performance can vary along a continuum of quality based on a variety of variables. One important variable is the human operator, who must complete specific tasks that are assessed in relation to a known standard. If the operator's performance exceeds that standard, it is labeled "acceptable," but, if the operator's performance fails to meet that standard, it is labeled "unacceptable."

1.4 Purpose

The purpose of the present study was threefold: (1) determine the reliability of participant ratings of controller performance obtained via the en route rating form; (2) determine the relationship between participant ratings and selected personality traits; and (3) further investigate the feasibility of using video and computerized graphical playback technology as a controller performance evaluation method.

1.4.1 Observing and Rating Behavior

SMEs evaluate performance. However, sometimes they apply their personal standard rather than the known standard. Personal standards are often influenced by the SME's experience, training, peer performance, and organizational standards (Anastasi, 1988). Real-time performance ratings must focus on concrete, observable behaviors. Even though the purpose of the rating should not influence the quality of the rating design or execution, it sometimes does.

Anastasi (1988) discussed using ratings as criterion measures for the verification of principally predictive indicators. The author stated that despite technical flaws and biases of evaluators, ratings are important sources of criterion information when they are collected under systematic conditions. She emphasized the importance of evaluator/rater training to increase reliability and validity while reducing common judgmental errors. Training can take many forms, but anything that heightens an evaluator's observational skills will probably improve rating quality, which affects reliability.

This study evaluated two types of reliability: inter-rater and intra-rater. Inter-rater reliability refers to the reliability of two or more independent raters. Intra-rater reliability refers to the reliability of an individual rater over time. Performance ratings can be sources of measurement error, so it is important to evaluate the consistency of such ratings. Inter-rater reliability is often evaluated through intra-class correlations, and researchers evaluate intra-rater reliability with Pearson's product moment correlations. Some standardized instruments have obtained reliabilities that are considered acceptable, with $r = .85$ or better (Gay, 1987, p. 141).

FAA researchers assess the reliabilities of many types of ratings, including over-the-shoulder (OTS) observational ratings. ATCSs have employed OTS observational ratings since the initiation of the ATC system. ATCSs believe they are qualified to observe and evaluate each other. However, a controversy exists over the value of observational performance ratings as compared to objective data that are obtained in the laboratory. One problem is that ATCSs are very decisive, and it can be hard to change their ideas about performance evaluation. When observing the same behavior at the same time under the same conditions, evaluators who have not been trained to systematically observe may produce different results from the trained evaluators. Under such circumstances, inter-rater reliability decreases.

OTS observational ratings have, however, often been used in ATC simulation research. Buckley, DeBaryshe, Hitchner, and Kohn (1983) included observational ratings in their performance evaluation system. Two observers completed performance evaluations every 10 minutes during the simulations. They used a 10-point scale to rate two areas: overall system effectiveness and individual controller judgment/technique. Inter-rater reliability ranged from .06 to .72.

1.4.2 Accommodating Subject Matter Experts

There are advantages and disadvantages to accommodating SMEs. The primary advantage is that they can make suggestions for changes to the ATCS performance-rating form that would increase its realism and its applicability to the field setting. Also, there is more participant buy-in possible. However, the corresponding disadvantage is that incorporating such suggestions may render the form facility- or use- specific. That is, if researchers incorporate SATCSs' suggestions into the form, and some of those suggestions apply only to the participant's particular facility, the form would be useless. The rating form used in this study was intended to be a research tool only, not to replace the evaluation form currently used in the field. Therefore, researchers included only those suggestions that related to observable behaviors that could be evaluated both by the form and in the research environment currently in use at the Research Development and Human Factors Laboratory (RDHFL). A related disadvantage is that SMEs bring to the research environment personal and facility biases that can influence the research process. When observing and evaluating controller performance as a group, the goal is for the SMEs to adopt mutual rating criteria in making their evaluations. If SMEs were using the same criteria in making their evaluations, researchers would be better able to assess the validity and usefulness of the rating form. SME biases should be addressed by including comments and suggested items in the form, but not if those items cannot be behaviorally evaluated.

2. Method

2.1 Participants

Nine SATCSs from five different air route traffic control centers (ARTCCs) participated in the present study. They ranged in age from 31 to 54 years (M = 44.56, SD = 7.45). The participants were full performance level SMEs with current experience in controlling traffic at their respective ARTCCs. They actively controlled traffic from 11 to 12 of the previous 12 months (M = 11.89, SD = 0.33). They had from 9 to 29 years experience controlling air traffic (M = 20.00, SD = 6.16), including from 1½ to 20 years experience training and evaluating controllers (M = 13.94, SD = 5.75). Finally, the participants had normal vision correctable to 20/30 with glasses.

2.2 Rating Form

The Sollenberger et al. (1997) TRACON rating form (see Appendix A) was the basis for developing the en route form. The TRACON rating form contained 24 items that assessed different areas of controller performance. They organized the performance areas into six categories, with an overall rating scale included for each category. Participants identified various observable behaviors that should be considered when assessing controller performance for each performance area. It contained an eight-point scale format, with statements describing the necessary controller actions for each scale. A comment section encouraging participants to write as much as possible appeared at the end of the form. This kept them oriented on controller behavior and helped to reduce their dependence on memory when assigning numerical ratings.

The en route rating form (see Appendix B) contained 26 items, including two 2-question items (items 15 and 19). However, participants concluded that they had insufficient information to rate items 13 and 18. The en route SATCSs gave significant input on organizing the rating form, and the researchers revised it according to their suggestions. They changed items 15 and 16 in the TRACON form to items 15A and 15B, added items 16 and 19B, and changed item 19 to item 19A. Further, the en route rating form provided space for comments after each item, with space for general comments at the end. Finally, as per technical instructions given to the researchers by the project technical lead, the N/A choice was eliminated from the rating scale in the en route form to discourage avoidance of an item. Instead, participants wrote N/A next to those items that they felt did not apply. The en route rating form included instructions on how to use the form and some assumptions about ATC and controller performance.

2.3 Airspace and Traffic Scenarios

The replay files used in the present study were recorded during a simulation study that investigated the effects of free flight conditions on controller performance, workload, and situation awareness (Endsley, Mogford, Allendoerfer, Snyder, & Stein, 1997). During that study, 10 controllers from the Jacksonville ARTCC (ZJX) worked traffic scenarios using the Greencove/Keystone sector, a combined high altitude sector.

Greencove/Keystone is responsible for altitudes of flight level (FL) 240 and higher and has four primary traffic flows. Southbound aircraft enter Greencove/Keystone from the northeast and northwest and continue south and southeast toward Fort Lauderdale, Miami, and West Palm Beach along the J45 or J79 airways. Aircraft are usually at their final altitude when they enter Greencove/Keystone. Some northbound aircraft leave Orlando International Airport and travel north or northwest along the J53 or J81 airways. They usually contact the sector at about FL 180 while climbing to an interim altitude of FL 230. They will be cleared to their final altitude when feasible. Other northbound aircraft depart from southeast Florida and enter Greencove/Keystone in the south, near Orlando. These aircraft continue north and northwest along the J53 and J81 airways. These aircraft are usually at their final altitude when they enter the sector but occasionally may need the controller to clear them to their final altitude. For Endsley et al.'s (1997) purposes, these aircraft were at their final altitude when they reached the sector.

The Greencove/Keystone sector is bordered below by the St. Augustine and St. Johns sectors, on the northeast by the States/Hunter combined sector, on the north-northwest by the Alma/Moultrie

combined sector, on the west by the Lake City/Ocala sector, on the southwest by the Mayo sector, on the south by the Miami ARTCC (ZMA) Boyel sector, and on the south-southeast by the ZMA Hobee sector. For Endsley et al.'s (1997) purposes, all adjacent sectors accepted all handoffs and approved all point-outs. Greencove/Keystone is bordered on the east by a warning area that is controlled by the US Navy. Civilian aircraft may enter the warning area only with special permission. For Endsley et al.'s purposes, the warning area was considered to be active, so no civilian aircraft were permitted to enter the area.

Endsley et al. (1997) used four types of scenarios. The present study incorporated only two of the four free flight study scenario types. The "condition A" scenarios included current ATC procedures. The "condition B" scenarios also utilized current ATC procedures but included direct routings.

2.4 Simulation Facility

Researchers conducted the present study in the briefing room of RDHFL at the FAA William J. Hughes Technical Center, Atlantic City International Airport, New Jersey. The RDHFL briefing room video projection system presented three views of the Endsley et al. (1997) study. The primary, center view was a graphical playback of the traffic scenario using the simulation software, ATCoach (UFA, Inc., 1992). The second view was recorded by a video camera located in a corner of the room in which the Endsley et al. study was conducted and showed an OTS view of the controller's upper body, workstation equipment, and radar display. The controller's head and arm movements and interactions with the workstation equipment were clearly visible, but it was not possible to read the writing on flight progress strips or the data on the radar display. The third view was a video recording of the simulation pilot's radar screen. All three views were simultaneously presented on different screens and synchronized with an audio recording of the controllers and simulation pilots.

2.5 Procedure

The study took 8 workdays. The first 4 days consisted of training and the last 4 days consisted of the actual replay evaluations. Participants completed several questionnaires during the first training session including the Background Questionnaire (see Appendix C for the questionnaire and Appendix D for their training and evaluation experience), the consent form regarding audio recording of discussions (see Appendix E), and the Sixteen Personality Factor (16PF) personality inventory. Participants completed the Final Questionnaire (see Appendix F) on the last evaluation day. On the Final Questionnaire, participants indicated the overall importance of the six performance areas to overall ATC performance. They selected a weight score between 0 and 100 for each area. The weights were to sum to 100. Higher weights indicated performance areas that the participants felt were more important to overall ATC performance.

The nine SATCSs participated as a single group in a 4-day training program in preparation for formal evaluations. The purpose of the training program was to teach participants to adopt common rating standards and educate them concerning the pitfalls of observation. A team of psychologists and SMEs conducted the training program in two separate sessions. The first

training session lasted 1 day and helped participants learn the airspace in the simulation. The second training session lasted 3 days and helped participants become proficient with the rating form.

In the first training session, researchers informed the participants about the goals of the study, how the study would be conducted, and what was expected from them as participants. They explained all aspects of the simulation setup, equipment, software, and data collection capabilities. A written description of the sector assisted participants in learning the airspace. The description included the Letters of Agreement (LOAs) and Standard Operating Procedures (SOPs) for the airspace and illustrated the sector layout and airways. The first session included several hands-on training scenarios during which participants had the opportunity to control some air traffic.

In the second training session, researchers explained the rating form design process and development work. They took several steps to encourage the participants to adopt common evaluation criteria for their ratings. First, the research team discussed some common rater biases and how to avoid them. The participants reviewed the rating form and discussed their interpretations of the terminology. Next, the participants used the rating form while viewing five practice scenarios. After each scenario, the participants discussed their ratings, what they saw in the scenario, and why they selected the ratings they did. Each discussion period lasted approximately 1 hour and helped to clarify any ambiguities in the rating form and identify any participant whose rating style differed a great deal from the others. Researchers modified the rating form in line with participants' input at the conclusion of the training program. The hourly schedules for training and evaluation activities are given in Appendix G.

After viewing the first practice replay, the participants requested a summary sheet, a copy of the rating form minus the space for comments after each item. They felt it would be easier to use the unified sheet and write comments on a sheet of scratch paper and then transcribe their comments onto the rating form itself. The researchers provided the participants with the requested form (see Appendix H) and asked the participants to attach their scratch paper to the rating form after the transcription of their comments.

For the evaluation phase of the study, the researchers selected replays from each of the 10 controllers who participated in the Endsley et al. (1997) study. As part of the design, the participants viewed four replays a second time to obtain a measure of intra-rater reliability. In total, the participants viewed 15 45-minute replays.

The presentation order of the scenarios ensured that similar ones were not viewed consecutively. Researchers organized the presentations so that only two of the controllers in the Endsley et al. (1997) study (controllers 1 and 5) were viewed twice before the last day of the study. However, controllers 1 and 5 were viewed on different days performing different scenarios than the first time they were viewed. Four scenarios, which had already been viewed once each, were viewed on the last evaluation day. These four scenarios provided the basis for examining ratings of repeated scenarios. The objective of this procedure was to minimize any carry-over effects between the replayed stimulus situations. The presentation order of the scenarios is shown in Appendix I.

In addition to the ratings obtained from the participants, the present study examined a set of SEMs routinely collected in ATC simulation research (Buckley et al., 1983). The participant ratings were compared to a subset of the SEMs, which included the number of conflict errors, controller assignments, controller transmissions, aircraft density, and controller workload. A list of the SEMs recorded during the present study is presented in Appendix J.

2.5.1 Replay Files

In preparing for the present study, the experimenters discovered some problems with the ATCoach replay files. The data from the Endsley et al. (1997) study were recorded in a version of ATCoach that accounted for controller entries such as interim altitudes, final altitudes, and movements of data blocks. However, those entries did not show up during the replay. In the replay files used in the present study, data blocks overlapped and incorrect interim and final altitudes were presented in the assigned altitude portion of the data block. Therefore, while the replays were running, two experienced simulation pilots made the necessary adjustments to the replay files to present a display more representative of the controller's actual planned view display. They used the computer to move data blocks and enter correct interim and final altitudes directly into the replay files. They did this the same way for each replay. The purpose of having simulation pilots make these adjustments was to prevent participants from negatively rating the controllers for not moving data blocks or not making correct altitude entries.

3. Results

The present study investigated

a. the reliability of participant ratings,

b. the relationship between participants' ratings and several SEMs,

c. the relationship between participants' ratings in the six overall performance areas, and

d. the relationship between participants' ratings and scores on the 16PF personality inventory.

3.1 Participant Ratings

The overall descriptive statistics for participant ratings of controller performance are presented in Table 1. However, participant ratings for items 13 and 18 on the en route rating form are not shown in Table 1 because the en route SATCSs did not rate the controllers on those items. The participants were unable to determine what the controllers were marking on the flight strips (shown on the OTS videotape), so they did not feel that they could adequately rate the controllers on item 13 (marking flight strips while performing other tasks). Also, the participants did not feel that they possessed adequate knowledge of the sector LOAs and SOPs, therefore, they felt they could not adequately rate the controllers on item 18 (showing knowledge of LOAs and SOPs).

Table 1. Participant Rating Grand Means

Item	Mean	SD
1. Maintaining Separation and Resolving Potential Conflicts	3.56	2.52
2. Sequencing Arrival, Departure, and En Route Aircraft Efficiently	4.86	2.07
3. Using Control Instructions Effectively/Efficiently	5.13	2.12
4. Overall Safe and Efficient Traffic Flow Scale Rating	4.01	2.17
5. Maintaining Situational Awareness	4.84	2.07
6. Ensuring Positive Control	4.35	2.05
7. Detecting Pilot Deviations from Control Instructions	5.36	1.82
8. Correcting Errors in a Timely Manner	5.33	1.79
9. Overall Attention and Situation Awareness Scale Rating	4.59	1.83
10. Taking Actions in an Appropriate Order of Importance	5.61	1.89
11. Preplanning Control Actions	5.55	1.93
12. Handling Control Tasks for Several Aircraft	5.56	1.77
14. Overall Prioritizing Scale Rating	5.34	1.83
15A. Providing Essential Air Traffic Control Information	3.36	2.08
15B. Providing Additional Air Traffic Control Information	4.04	2.12
16. Providing Coordination	3.90	2.33
17. Overall Providing Control Information Scale Rating	3.63	2.02
19A. Showing Knowledge of Aircraft Capabilities and Limitations	5.36	1.97
19B. Showing Effective Use of Equipment	5.58	1.84
20. Overall Technical Knowledge Scale Rating	5.19	1.91
21. Using Proper Phraseology	4.75	2.06
22. Communicating Clearly and Efficiently	5.40	2.07
23. Listening to Pilot Readbacks and Requests	5.07	1.90
24. Overall Communicating Scale Rating	4.91	1.93

3.1.1 Inter-Rater Reliability of Participant Ratings

The intra-class correlation assessed inter-rater reliability for each item of the en route rating form. These correlations are presented in Table 2. Items 13 and 18 are not shown in Table 2 because the en route participants did not rate the controllers on those items.

Table 2. Inter-Rater Reliability for the En Route Rating Form

Item	Inter-Rater Reliability
1. Maintaining Separation and Resolving Potential Conflicts	.74
2. Sequencing Arrival, Departure, and En Route Aircraft Efficiently	.40
3. Using Control Instructions Effectively/Efficiently	.47
4. Overall Safe and Efficient Traffic Flow Scale Rating	.72
5. Maintaining Situational Awareness	.60
6. Ensuring Positive Control	.45
7. Detecting Pilot Deviations from Control Instructions	.65
8. Correcting Errors in a Timely Manner	.61
9. Overall Attention and Situation Awareness Scale Rating	.61
10. Taking Actions in an Appropriate Order of Importance	.64
11. Preplanning Control Actions	.56
12. Handling Control Tasks for Several Aircraft	.61
14. Overall Prioritizing Scale Rating	.66
15A. Providing Essential Air Traffic Control Information	.53
15B. Providing Additional Air Traffic Control Information	.47
16. Providing Coordination	.62
17. Overall Providing Control Information Scale Rating	.55
19A. Showing Knowledge of Aircraft Capabilities and Limitations	.27
19B. Showing Effective Use of Equipment	.35
20. Overall Technical Knowledge Scale Rating	.40
21. Using Proper Phraseology	.47
22. Communicating Clearly and Efficiently	.56
23. Listening to Pilot Readbacks and Requests	.47
24. Overall Communicating Scale Rating	.51
Weighted Overall Performance Score	.65

The reliability coefficients of the scales included in the en route rating form ranged from $r = .27$ to $r = .74$. Thirty-five percent of the coefficients exceeded $r = .60$ and 8% exceeded $r = .70$. The overall ratings for each performance category were generally more reliable than the individual ratings included in each category. The weighted overall performance score was $r = .65$. The weighted overall performance score was calculated by using the weighting values that indicated the relative importance of the six performance categories included in the en route rating form. Participants provided these weighting values on the Final Questionnaire (see Section 2.5, Procedure). Specifically, the weight for each category was multiplied by the mean rating for each

category (the mean of the ratings for each evaluation item within a category). The results were summed to produce a weighted overall performance score ranging from 1.0 to 8.0.

Because two of the four types of Endsley et al. (1997) scenarios were used in the present study, researchers calculated the intra-class correlations for both types of scenarios. The condition A scenarios included current ATC procedures. The condition B scenarios also utilized current ATC procedures but included direct routings. The intra-class correlations for the two conditions are presented in Table 3. Items 13 and 18 are not shown in Table 3 because the en route

Table 3. Inter-Rater Reliability for Endsley et al. (1997) Condition A and Condition B Scenarios

Item	Condition A	Condition B
1. Maintaining Separation and Resolving Potential Conflicts	.49	.85
2. Sequencing Arrival, Departure, and En Route Aircraft Efficiently	.24	.42
3. Using Control Instructions Effectively/Efficiently	.49	.34
4. Overall Safe and Efficient Traffic Flow Scale Rating	.53	.72
5. Maintaining Situational Awareness	.30	.68
6. Ensuring Positive Control	.29	.52
7. Detecting Pilot Deviations from Control Instructions	.19	.75
8. Correcting Errors in a Timely Manner	.33	.76
9. Overall Attention and Situation Awareness Scale Rating	.45	.66
10. Taking Actions in an Appropriate Order of Importance	.55	.64
11. Preplanning Control Actions	.38	.60
12. Handling Control Tasks for Several Aircraft	.53	.56
14. Overall Prioritizing Scale Rating	.60	.58
15A. Providing Essential Air Traffic Control Information	.53	.45
15B. Providing Additional Air Traffic Control Information	.38	.52
16. Providing Coordination	.56	.51
17. Overall Providing Control Information Scale Rating	.49	.50
19A. Showing Knowledge of Aircraft Capabilities and Limitations	.29	.17
19B. Showing Effective Use of Equipment	.35	.41
20. Overall Technical Knowledge Scale Rating	.33	.35
21. Using Proper Phraseology	.60	.29
22. Communicating Clearly and Efficiently	.65	.46
23. Listening to Pilot Readbacks and Requests	.33	.59
24. Overall Communicating Scale Rating	.38	.48
Overall Weighted Performance Score	.10	.67

participants did not rate the controllers on those items. As can be seen in Table 3, the reliability of the condition A scenarios was often lower than that for the condition B scenarios. In condition A, items 3, 14, 15A, 16, 19A, 21, and 22 were the only items whose reliability was greater than that of their condition B counterparts.

3.1.2 Intra-Rater Reliability of Participant Ratings

Researchers computed Pearson's product moment correlations to evaluate intra-rater reliability on four repeated Endsley et al. (1997) scenarios, two of which were condition A scenarios and two of which were condition B scenarios. These correlations are presented in Table 4. As can be

Table 4. Intra-Rater Reliability for the En Route Rating Form

Item	Intra-Rater Reliability
1. Maintaining Separation and Resolving Potential Conflicts	.69
2. Sequencing Arrival, Departure, and En Route Aircraft Efficiently	.75
3. Using Control Instructions Effectively/Efficiently	.57
4. Overall Safe and Efficient Traffic Flow Scale Rating	.84
5. Maintaining Situational Awareness	.70
6. Ensuring Positive Control	.69
7. Detecting Pilot Deviations from Control Instructions	.38
8. Correcting Errors in a Timely Manner	.51
9. Overall Attention and Situation Awareness Scale Rating	.75
10. Taking Actions in an Appropriate Order of Importance	.67
11. Preplanning Control Actions	.74
12. Handling Control Tasks for Several Aircraft	.70
14. Overall Prioritizing Scale Rating	.77
15A. Providing Essential Air Traffic Control Information	.73
15B. Providing Additional Air Traffic Control Information	.79
16. Providing Coordination	.73
17. Overall Providing Control Information Scale Rating	.81
19A. Showing Knowledge of Aircraft Capabilities and Limitations	.55
19B. Showing Effective Use of Equipment	.55
20. Overall Technical Knowledge Scale Rating	.65
21. Using Proper Phraseology	.74
22. Communicating Clearly and Efficiently	.87
23. Listening to Pilot Readbacks and Requests	.51
24. Overall Communicating Scale Rating	.79

Item	Intra-Rater Reliability
Overall Weighted Performance Score	.87

seen in Table 4, the reliability coefficients of the scales included in the en route rating form, which ranged from r = .38 to r = .87, were somewhat higher than the inter-rater coefficients. Sixty-three percent of the coefficients exceeded r = .60 and 13% exceeded r = .80. The overall ratings for each performance category were generally more reliable than the individual ratings within each category. The overall weighted performance score was r = .87. The intra-rater reliability of the overall weighted performance score was greater than the inter-rater reliability of the overall weighted performance score, with r = .65 (inter-rater) vs. r = .87 (intra-rater). Items 13 and 18 are not shown in Table 4 because the en route participants did not rate the controllers on those items.

3.2 Relationship Between Participant Ratings and System Effectiveness Measures

Endsley et al. (1997) collected SEMs, objective measures of controller performance. A correlation analysis determined the relationship between participant performance ratings and the SEMs. Only two SEMs correlated significantly with participant performance ratings, number of speed assignments (r = -.46) and number of ground-to-air transmissions (r = .28). Table 5 presents the descriptive statistics for the SEMs.

Table 5. Performance Measures (SEMs)

SEM	Mean (Freq.)	SD
NCNF (Number of en route conflicts)	0.38	0.49
NALT (Number of altitude changes)	35.13	12.86
NHDG (Number of heading changes)	26.63	11.39
NSPD (Number if airspeed changes)	1.13	0.79
NPTT (Number of push-to-talk communications)	74.50	38.95
CMAV (Cumulative Average of System Activity/Aircraft Density)	1.67	0.66
ATWIT (Air Traffic Workload Input Technique)	3.25	1.96

3.3 Intercorrelations Among Overall Performance Area Ratings

A correlation analysis determined the relationship between the participant ratings in the six overall performance areas: Maintaining Safe and Efficient Traffic Flow, Maintaining Attention and Situation Awareness, Prioritizing, Providing Control Information, Technical Knowledge, and Communication. As can be seen in Table 6, the intercorrelations among the six overall performance areas ranged from r = .55 to r = .80. All of these correlations were significant, $p < .01$.

The relationships between the six "overall" scale ratings and the ratings of their corresponding subscales were also analyzed. The correlations ranged from r = .72 to r = .94. All of the

"overall" scale ratings correlated significantly with their corresponding subscale ratings at the $p < .01$ level. This indicates expected redundancy between subscale ratings and their corresponding "overall" scale ratings.

Table 6. Intercorrelations Among the Overall Performance Areas

	1	2	3	4	5	6
1. Overall Safe and Efficient Traffic Flow Scale Rating						
2. Overall Attention and Situational Awareness Scale Rating	.80	.				
3. Overall Prioritizing Scale Rating	.73	.77	.			
4. Overall Providing Control Information Scale Rating	.79	.71	.68			
5. Overall Technical Knowledge Scale Rating	.70	.71	.71	.76	.	
6. Overall Communication Scale Rating	.55	.66	.66	.62	.65	

3.4 Relationship Between Participant Ratings and Scores on the 16PF Personality Inventory

The researchers evaluated whether rater background related to the ratings assigned. The participants completed the 16PF personality inventory during the first day of training. The Institute for Personality and Ability Testing (IPAT) scored the responses and returned the results as standardized scores.

Data included how the participants scored for the five global factors (i.e., extroversion, anxiety, tough-mindedness, independence, and self-control). A second information set indicated how the participants scored for the 16 basic factors (warmth, reasoning, emotional stability, dominance, liveliness, rule conscious, social boldness, sensitivity, vigilance, abstractness, privateness, apprehension, openness to change, self reliance, perfectionism, and tension). The descriptive statistics for how participants scored in terms of the 16PF global and basic factors are presented in Appendix K.

The researchers correlated the sten scores from both the global factors and basic factors with participant ratings of controller performance. The results of the correlational analysis on the global factors are presented in Appendix L, and the results of the correlational analysis on the basic factors are presented in Appendix M.

As can be seen in Appendix L, the correlations between participant ratings and the 16PF global factors ranged from r = -.49 to r = .48. These relationships were not strong. However, some were significant from zero. Extroversion was significantly correlated with 13 of the 26 scales, anxiety was significantly correlated with six scales, and self-control was significantly correlated with 18 scales. Additionally, these variables were significantly correlated with the overall weighted performance score. Tough-mindedness was significantly correlated with nine scales, and independence was significantly correlated with nine scales. However, these two variables were not significantly correlated with the overall weighted performance score. A p value of .05 was used as the criterion of significance.

Of the 16 basic factors, the researchers were primarily interested in reasoning, rule conscious, vigilance, openness to change, perfectionism, and tension because a controller must be able to develop a timely solution to a problem. The controller must be able to follow rules (LOAs and SOPs) and be vigilant when monitoring the radar screen. Researchers asked the observers to change their normal way of thinking about the rating process by adopting mutual evaluation

criteria and not following personal or ARTCC biases. Controllers tend to have very high standards (FAA, 1998). Appendix M shows the correlations for these six factors ranged from r = -.45 to r = .43.

Rule conscious was significantly correlated with 16 scales and perfectionism was significantly correlated with 14 scales. Both variables were significantly correlated with the overall weighted performance score. Openness to change was significantly correlated with eight scales. However, tension only was significantly correlated with listening to pilot readbacks and requests, and reasoning only was significantly correlated with preplanning control actions. A p value of .05 was used as the criterion of significance.

3.5 Summary of Final Questionnaire

On the Final Questionnaire, participants indicated the overall importance of the six performance areas to overall ATC performance. They selected a weight score between 0 and 100 for each area. These weights summed to 100. Higher weights indicated performance areas that the participants felt were more important to overall ATC performance. Table 7 presents the mean weights that participants assigned to each of the six major categories.

Table 7. Mean Weights Assigned to Each Performance Category

	Maintaining a Safe and Efficient Traffic Flow	Maintaining Attention and Situation Awareness	Communicating	Providing Control Information	Technical Knowledge	Prioritizing
Mean	36.67	20.56	13.33	11.11	10.00	9.44
SD	5.59	4.64	3.54	3.33	3.54	4.64

Participants were also asked to rate both the radar presentation and the training period on a scale of 1 (indicating poor quality) to 10 (indicating excellent quality). The participant rating of the radar display was M = 3.44, SD = 2.13 and the participant rating of the training period was M = 7.22, SD = 1.64.

4. Discussion

4.1 Reliability of Participant Ratings

Participants were relatively consistent over time with their own ratings (intra-rater reliability). Results on the agreement between raters (inter-rater reliability), however, were somewhat disappointing. The low inter-rater reliability may have been caused by several factors including the simulation replay problems that occurred, artifacts of the en route environment, the differences in the types of scenarios viewed (condition A or condition B), or an overall lack of variability. Further, intra-class correlations used to assess the inter-rater reliability tend to be lower than Pearson bivariate correlations used to measure within-rater agreement. The intra-rater reliability of the overall weighted performance score was greater than the inter-rater reliability of the overall weighted performance score: r = .65 (inter-rater) vs. r = .87 (intra-rater). This is not an

uncommon finding and reflects the difficulty of having professionals, who may be internally consistent with their standards for performance, try to come to a common frame of reference. Simulation replay artifacts were most likely the culprits of the low inter-rater reliability.

4.1.1 Inter-Rater Reliability

There are several possible reasons for low inter-rater reliability. First, events that occurred in the training session may have influenced the reliability of the en route rating form. Second, the specific changes made to the form that were recommended by the en route participants could have also influenced reliability.

The Endsley et al. (1997) condition A scenarios were less reliable in terms of inter-rater reliability, than the condition B scenarios. Controllers had to take more control actions in the condition A scenarios (which investigated controller performance under current ATC procedures) than they did in the condition B scenarios (which investigated controller performance under direct routings, otherwise current ATC procedures). It appears that there was more agreement on performance ratings when controllers had to take fewer control actions.

4.1.2 Intra-Rater Reliability

The results of this study indicate that participants retained the same evaluation criteria when viewing scenarios more than once. The intra-rater reliabilities indicated that they observed similar events and evaluated them the same way each time they observed a specific scenario. It would not be unreasonable for participants to view different things on repeated viewings of a scenario because they could have missed something the first time. If the intra-rater reliabilities were particularly low, that would indicate that participants might not have viewed the same events on repeated viewings of scenarios. This would most likely be due to participants changing evaluation criteria between scenario viewings.

The results of this study indicate, however, that participants did maintain performance evaluation criteria between repeated viewings of scenarios. In many cases, the intra-rater reliabilities varied across scales but not greatly, given the complexity of the tasks.

4.2 Relationship Between Participant Ratings and System Effectiveness Measures

Of the six Endsley et al. (1997) SEMs evaluated in the present study, only two were significantly correlated with the weighted overall performance score: number of speed assignments (NSPD) and number of ground-to-air transmissions (NPTT). NSPD was inversely correlated with the weighted overall performance score. Thus, the lower the number of speed assignments the controllers made, the higher their weighted overall performance scores. This would indicate that the participants rated the more efficient controllers (those who made fewer speed assignments or those who took fewer control actions) higher. That is how controllers are usually evaluated in the field: the more efficient the controller is, the higher the performance rating the controller receives.

4.3 Intercorrelations Among Overall Performance Area Ratings

The relationships between observer ratings in the six overall performance areas indicate some redundancy across areas. The overall scale ratings were all significantly correlated with each

other at $p < .01$. The overall scale ratings were also all significantly correlated with their respective subscales at $p < .01$. Although the correlations were not perfect, this does indicate some redundancy between the overall scales and their respective subscales. Thus, the same results would probably have been achieved if the subscales were not included in the rating form.

This redundancy can also be viewed as internal consistency. There is sufficient consistency across categories, but they may not be redundant. However, because the correlations are not perfect, there is also some unique variance within each area.

4.4 Relationship Between Participant Ratings and Scores on the 16PF Personality Inventory

Personality is a difficult construct to measure. The 16PF personality inventory is one of the most widely used and researched instruments available. Two-week test-retest reliabilities of the 16PF fifth edition ranged from $r = .69$ to $r = .87$, with a median of $r = .80$. Two-month test-retest reliabilities of the 16PF fifth edition ranged from $r = .56$ to $r = .79$, with a median of $r = .69$ (Conn & Rieke, 1994). Thus, the 16PF is considered to be a reliable measure of normal personality. The 16PF results of the present study indicate that what controllers bring with them to an observational rating environment in terms of personality characteristics, does matter. They suggest that dimensions that are stable parts of who the participants are may relate to some of the inter-rater reliability issues. These dimensions are difficult to overcome with only 1 week of training.

Each of the 16PF global factors was significantly correlated with several of the scales included in the en route rating form. Thus, participants' personality characteristics were related to their ratings of controller performance. Four scales (sequencing arrival, departure, and en route aircraft efficiently; ensuring positive control; listening to pilot readbacks and requests; and overall communicating scale rating) were significantly correlated with all five 16PF global factors. These results are not surprising. Controllers must properly sequence aircraft, maintain attention and situational awareness, preplan and prioritize their control actions, provide essential ATC information to pilots, and communicate with pilots and other controllers efficiently in order to perform adequately and prevent operational errors from occurring. The ideal ATCS, the person who would be best suited to perform these tasks, would be more extroverted, less anxious (able to stay cool in tough situations), more tough-minded (does not change his/her mind very easily and is not indecisive), independent (confident and able to stand his/her ground), and in control (not get excited and lose control of the situation).

Of the 16PF basic factors, the researchers were primarily interested in how reasoning, rule conscious, vigilance, openness to change, perfectionism, and tension related to participant rating of controller performance. That is because controllers must be able to reason how to prevent an operational error or problem from occurring or, if one does occur, the controller must be able to reason a solution to the problem. The controller must follow rules (LOAs, SOPs, and FAA 7110.65L [FAA, 1998]). The controller must remain vigilant when monitoring the radar screen. The participants in the present study were asked to change their normal ways of rating controller performance (they were asked to ignore personal and/or facility biases and to adopt mutual rating standards), so they needed to be open to change. ATC is strictly regulated (FAA, 1998). There is very little room for error because people's lives depend on both pilots and controllers, so the successful ATCS must be a perfectionist. The successful ATCS must also be able to tolerate

tension because the world of ATC is filled with tension. Thus, the successful ATCS would be able to reason solutions to problems, be conscious of ATC rules, be vigilant, be open to change, be perfectionistic, and be able to tolerate high tension. Several of the scales included in the revised rating form were significantly correlated with at least three (half) of these factors, including the following: sequencing arrival, departure, and en route aircraft efficiently; ensuring positive control; overall attention and situation awareness scale rating; providing essential ATC information; listening to pilot readbacks and requests; and overall communicating scale rating.

What do significant correlations between 16PF scores and performance ratings really mean? The 16PF has a good reputation in academia, research, and the clinical field. The products generated are generally accepted as meaningful indicators of the respondent's personality, a relatively enduring set of traits reflecting who the respondent is. The fact that significant correlations exist suggests that part of the variance in ratings may be related to participant background. These correlations are far from perfect. This indicates that there is considerable variance with which participant personality does correlate.

4.5 Summary of Final Questionnaire

The primary goal of ATC is to maintain safety. Participants gave Maintaining a Safe and Efficient Traffic Flow the highest priority when rating controller performance. Thus, raters felt that this was the most important of the controllers' many tasks.

Participants did not feel that the radar display showed sufficient information for them to make their evaluations, as compared to viewing controllers "live." Participants also rated the training period on a scale of 1 to 10. The participants did, however, feel that the training period was sufficient for them to become familiar with the rating form.

In general, participants did not believe that, compared to viewing controllers live, the radar display showed sufficient information for them to make their evaluations. One reason was that the simulation pilot videotape was unclear. Participants were generally unable to distinguish letters and numbers. With all of the problems with the simulation pilot videotape considered, participants did feel that the pilot radar was the best source to acquire accurate altitude and route information. However, from viewing the simulation pilot's radar videotape, participants were unable to determine what the controller was doing (e.g., dropping data blocks, entering interim altitudes, taking hand-offs, and making point-outs).

During some scenarios, the sound and the radar tape were not completely synchronized, which was annoying. Participants felt that the audio feature could be good, but only if it was synchronized with the pseudo-pilot's radar screen and/or the OTS videotape.

Participants indicated that the OTS videotape was not very useful. They could not see the strips well enough to determine if controllers were marking them correctly. Participants felt that the OTS videotape was useful only to determine if controllers were looking at the scope or "joking in the aisles."

The problems with ATCoach also seem to have affected participants' evaluations. Participants were generally unable to determine if these problems were related to ATCoach or to the controller. They suggested that ATCoach should be more realistic and were generally annoyed with it. In its present state, the simulation software was a distraction to performing effective evaluations.

The participants felt that the ATCSs whom they viewed on tape performed poorly. They felt that the controllers should have approached the study as if they were controlling a real life sector. A participant stated that they became so angry at the controllers' performance, they felt their evaluations were becoming harsher as time went on.

One participant had a lot to say about the evaluation form. The participant felt that the form was much too cumbersome as a usable, in-field document. The evaluation form used in this study, however, was never intended to be an in-field document. It was intended to be a research tool and requires that participants write extensively, which they do not do in the field. They depend heavily on memory when doing periodic and recertification ratings.

5. Conclusions

The present study evaluated the reliability of the revised rating form using en route participants who observed, via videotape and computerized graphical replay, controllers performing an en route simulation. The low inter-rater reliability of the en route rating form may have been caused by ATCoach problems that interrupted the study numerous times. Researchers instructed the participants to ignore the system problems to the limit of their collective abilities. However, even though they did their best to comply, some adverse impact may have occurred.

This study, despite its problems, did lead to some viable conclusions. Intra-rater reliability was greater than inter-rater reliability. Supervisory controllers came to the participant-rating task with personal and facility backgrounds, which can influence results.

Performance evaluation is an inherently complex process. There will never be a perfect OTS evaluation form or training process. However, subjective rating has been a mainstay in aviation and will continue. Researchers will likewise continue to try and improve the process, its reliability, and subsequent validity.

6. Recommendations

Some recommendations follow from the previous conclusions. The researchers should conduct another video evaluation, using the same rating form but computerized graphical replays and videotapes from an en route study other than the Endsley et al. (1997) study. This future study will enable researchers to determine whether or not the unacceptable low inter-rater reliability was due to ATCoach problems.

The en route rating form has potential as an assessment tool that provides data about controller performance. These data are not available from any other source. Future research efforts will focus on identifying the sources of measurement error and making whatever changes are necessary to produce a reliable performance assessment tool. Future development and evaluation of the rating form will continue and will improve the performance evaluation process.

References

Anastasi, A. (1988). *Psychological testing (6th Edition)*. New York: Macmillan Publishing Company.

Buckley, E. P., DeBaryshe, B. D., Hitchner, N., & Kohn, P. (1983). *Methods and measurements in real-time air traffic control system simulation* (DOT/FAA/CT-TN83/26). Atlantic City, NJ: DOT/FAA Technical Center.

Conn, S. T. & Rieke, M. L. (Eds.) (1994). *16PF fifth edition technical manual*. Champaign, Illinois: Institute for Personality and Ability Testing, Inc.

Endsley, M., Mogford, R., Allendoerfer, K., Snyder, M., & Stein, E. S. (1997). *Effects of free flight conditions on controller performance, workload, and situation awareness* (DOT/FAA/CT-TN97/12). Atlantic City, NJ: DOT/FAA Technical Center.

Federal Aviation Administration. (1998). *Order 7110.65L: Air traffic control*. Washington, DC: US Department of Transportation.

Gay, L. R. (1987*). Educational research: Competencies for analysis and application (3rd edition)*. Columbus: Merrill Publishing Company.

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

UFA, Inc. (1992). ATCoach [Computer software]. Lexington, MA: Authors.

Appendix A

Observer Rating Form -- TRACON

OBSERVER RATING FORM

Observer Code _____ Date _____

Controller 1 2 3 4 Sector JAX GEN Traffic LO HI

INSTRUCTIONS

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

Rating	Scale Point Description
1	Controller demonstrated *extremely* poor judgment in making control decisions and *very* frequently made errors
2	Controller demonstrated poor judgment in making some control decisions and occasionally made errors
3	Controller made questionable control decisions using poor control techniques which led to restricting the normal traffic flow
4	Controller demonstrated the ability to keep aircraft separated but used spacing and separation criteria which was excessive
5	Controller demonstrated *adequate* judgment in making control decisions
6	Controller demonstrated *good* judgment in making control decisions using efficient control techniques
7	Controller *frequently* demonstrated *excellent* judgment in making control decisions using extremely good control techniques
8	Controller *always* demonstrated excellent judgment in making even the most difficult control decisions while using outstanding control techniques
NA	Not Applicable - There was not an opportunity to observe performance in this particular area during the simulation

OBSERVER RATING FORM

(continued)

I - MAINTAINING SAFE AND EFFICIENT TRAFFIC FLOW

1. Maintaining Separation and Resolving Potential Conflicts ..... 1 2 3 4 5 6 7 8 NA
   - using control instructions that maintain safe aircraft separation
   - detecting and resolving impending conflicts early
2. Sequencing Arrival and Departure Aircraft Efficiently..... 1 2 3 4 5 6 7 8 NA
   - using efficient and orderly spacing techniques for arrival and departure aircraft
   - maintaining safe arrival and departure intervals that minimize delays
3. Using Control Instructions Effectively ..... 1 2 3 4 5 6 7 8 NA
   - providing accurate navigational assistance to pilots
   - avoiding clearances that result in the need for additional instructions to handle aircraft completely
   - avoiding excessive vectoring or over-controlling
4. Overall Safe and Efficient Traffic Flow Scale Rating ..... 1 2 3 4 5 6 7 8 NA

II - MAINTAINING ATTENTION AND SITUATION AWARENESS

5. Maintaining Awareness of Aircraft Positions ..... 1 2 3 4 5 6 7 8 NA
   - avoiding fixation on one area of the radar scope when other areas need attention
   - using scanning patterns that monitor all aircraft on the radar scope
6. Ensuring Positive Control ..... 1 2 3 4 5 6 7 8 NA
7. Detecting Pilot Deviations from Control Instructions ..... 1 2 3 4 5 6 7 8 NA
   - ensuring that pilots follow assigned clearances correctly
   - correcting pilot deviations in a timely manner
8. Correcting Own Errors in a Timely Manner..... 1 2 3 4 5 6 7 8 NA
9. Overall Attention and Situation Awareness Scale Rating ..... 1 2 3 4 5 6 7 8 NA

III - PRIORITIZING

10. Taking Actions in an Appropriate Order of Importance 1 2 3 4 5 6 7 8 NA
    - resolving situations that need immediate attention before handling low priority tasks
    - issuing control instructions in a prioritized, structured, and timely manner
11. Preplanning Control Actions 1 2 3 4 5 6 7 8 NA
    - scanning adjacent sectors to plan for inbound traffic
    - studying pending flight strips in bay
12. Handling Control Tasks for Several Aircraft 1 2 3 4 5 6 7 8 NA
    - shifting control tasks between several aircraft when necessary
    - avoiding delays in communications while thinking or planning control actions
13. Marking Flight Strips while Performing Other Tasks 1 2 3 4 5 6 7 8 NA
    - marking flight strips accurately while talking or performing other tasks
    - keeping flight strips current
14. Overall Prioritizing Scale Rating 1 2 3 4 5 6 7 8 NA

OBSERVER RATING FORM

(continued)

IV - PROVIDING CONTROL INFORMATION

15. Providing Essential Air Traffic Control Information 1 2 3 4 5 6 7 8 NA
    - providing mandatory services and advisories to pilots in a timely manner
    - exchanging essential information
16. Providing Additional Air Traffic Control Information 1 2 3 4 5 6 7 8 NA
    - providing additional services when workload is not a factor
    - exchanging additional information
17. Overall Providing Control Information Scale Rating 1 2 3 4 5 6 7 8 NA

V - TECHNICAL KNOWLEDGE

18. Showing Knowledge of LOAs and SOPs 1 2 3 4 5 6 7 8 NA
    - controlling traffic as depicted in current LOAs and SOPs
    - performing hand-off procedures correctly
19. Showing Knowledge of Aircraft Capabilities and Limitations 1 2 3 4 5 6 7 8 NA
    - avoiding clearances that are beyond aircraft performance parameters
    - recognizing the need for speed restrictions and wake turbulence separation
20. Overall Technical Knowledge Scale Rating 1 2 3 4 5 6 7 8 NA

VI - COMMUNICATING

21. Using Proper Phraseology 1 2 3 4 5 6 7 8 NA
    - using words and phrases specified in ATP 7110.65
    - using ATP phraseology that is appropriate for the situation
    - avoiding the use of excessive verbiage
22. Communicating Clearly and Efficiently 1 2 3 4 5 6 7 8 NA
    - speaking at the proper volume and rate for pilots to understand
    - speaking fluently while scanning or performing other tasks
    - clearance delivery is complete, correct and timely
    - providing complete information in each clearance
23. Listening to Pilot Readbacks and Requests 1 2 3 4 5 6 7 8 NA
    - correcting pilot readback errors
    - acknowledging pilot or other controller requests promptly
    - processing requests correctly in a timely manner
24. Overall Communicating Scale Rating 1 2 3 4 5 6 7 8 NA

OBSERVER RATING FORM
(continued)

I - MAINTAINING SAFE AND EFFICIENT TRAFFIC FLOW

1. Maintaining Separation and Resolving Potential Conflicts

2. Sequencing Arrival and Departure Aircraft Efficiently

3. Using Control Instructions Effectively

4. Other Actions Observed in Safe and Efficient Traffic Flow

II - MAINTAINING ATTENTION AND SITUATION AWARENESS

5. Maintaining Awareness of Aircraft Positions

6. Ensuring Positive Control

7. Detecting Pilot Deviations from Control Instructions

8. Correcting Own Errors in a Timely Manner

9. Other Actions Observed in Attention and Situation Awareness

OBSERVER RATING FORM
(continued)

III - PRIORITIZING

10. Taking Actions in an Appropriate Order of Importance

11. Preplanning Control Actions

12. Handling Control Tasks for Several Aircraft

13. Marking Flight Strips while Performing Other Tasks

14. Other Actions Observed in Prioritizing

IV - PROVIDING CONTROL INFORMATION

15. Providing Essential Air Traffic Control Information

16. Providing Additional Air Traffic Control Information

17. Other Actions Observed in Providing Control Information

OBSERVER RATING FORM
(continued)

V - TECHNICAL KNOWLEDGE

18. Showing Knowledge of LOAs and SOPs

19. Showing Knowledge of Aircraft Capabilities and Limitations

20. Other Actions Observed in Technical Knowledge

VI - COMMUNICATING

21. Using Proper Phraseology

22. Communicating Clearly and Efficiently

23. Listening to Pilot Readbacks and Requests

24. Other Actions Observed in Communicating

Appendix B

Participant Rating Form -- En Route

Observer Code _____ Date _____

Participant: 1 2 3 4 5 6 7 8 9 10

Condition: A B Scenario: 1 2 3 4 5 6 7 8 9 10 11 12 13 14

INSTRUCTIONS

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

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

Rating scale descriptors
Remove this Page and keep it available while doing ratings

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

I - MAINTAINING SAFE AND EFFICIENT TRAFFIC FLOW

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

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

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

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

II - MAINTAINING ATTENTION AND SITUATIONAL AWARENESS

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

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

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

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

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

III - PRIORITIZING

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

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

12. Handling Control Tasks for Several Aircraft **1 2 3 4 5 6 7 8**

- shifting control tasks between several aircraft when necessary
- communicating in timely fashion while sharing time with other actions

13. Marking Flight Strips while Performing Other Tasks..... **1 2 3 4 5 6 7 8**

- marking flight strips accurately while talking or performing other tasks
- keeping flight strips current

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

IV - PROVIDING CONTROL INFORMATION

15a. Providing Essential Air Traffic Control Information..... **1 2 3 4 5 6 7 8**

- providing mandatory services and advisories to pilots in a timely manner
- exchanging essential information

15b. Providing Additional Air Traffic Control Information **1 2 3 4 5 6 7 8**

- providing additional services when workload is not a factor
- exchanging additional information

16. Providing Coordination..... **1 2 3 4 5 6 7 8**

- providing effective coordination
- providing timely coordination
- using proper point-out procedures
- performing hand-off procedures properly

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

V - TECHNICAL KNOWLEDGE

18. Showing Knowledge of LOAs and SOPs **1 2 3 4 5 6 7 8**
- controlling traffic as depicted in current LOAs
- controlling traffic as depicted in current SOPs

19a. Showing Knowledge of Aircraft Capabilities and Limitations..... **1 2 3 4 5 6 7 8**
- using appropriate speed, vectoring, and/or altitude assignments to separate aircraft with varied flight capabilities
- issuing clearances that are within aircraft performance parameters

19b. Showing Effective Use of Equipment..... **1 2 3 4 5 6 7 8**
- updating of data blocks
- using equipment capabilities

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

VI – COMMUNICATING

21. Using Proper Phraseology..... **1 2 3 4 5 6 7 8**

- using words and phrases specified in the 7110.65
- using phraseology that is appropriate for the situation
- using minimum necessary verbiage

22. Communicating Clearly and Efficiently..... **1 2 3 4 5 6 7 8**

- speaking at the proper volume and rate for pilots to understand
- speaking fluently while scanning or performing other tasks
- ensuring clearance delivery is complete
- speaking with confident, authoritative tone of voice

23. Listening to Pilot Readbacks and Requests..... **1 2 3 4 5 6 7 8**

- correcting pilot readback errors
- acknowledging pilot or other controller requests promptly

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

GENERAL COMMENTS

Appendix C

BACKGROUND QUESTIONNAIRE

Observer Code _____ Date _____

INSTRUCTIONS

This questionnaire is designed to obtain information about your background as an air traffic control specialist. The information will be used to describe the participants in this study as a group in written or oral reports. Your identity will remain anonymous, so do not write your name on the form. Instead, your data will be identified by an observer code known only to yourself and the researchers conducting this study.

---

1. What is your job position or title?

   _____

2. What is your age?

   _____ years

3. How many years have you worked as an air traffic control specialist?

   _____ years

4. How many of the past 12 months have you actively controlled traffic?

   _____ months

5. How many years of experience do you have training and evaluating air traffic controllers?

   _____ years

6. Please briefly describe your air traffic control training and evaluation experience.

   _____

   _____

   _____

   _____

   _____

   _____

   _____

Appendix D

Participants' Air Traffic Control Training and Evaluation Experience

OBSRVR	AIR TRAFFIC CONTROL TRAINING AND EVALUATION EXPERIENCE
1	For the past 1 1/2 years I have been the Training Supervisor for my area. I am the liaison between the Training Department and Area 2. For the last 6 months I have been deeply involved in remedial training and performance improvement of an FPL who was lacking basic skills.
2	OJTI -- Oceanic control, ARTCC Academy Instructor -- Screen, OJTI Cadre course Field Instructor -- OJTI class Trained in 4 field facilities
3	OJTI -- 1987-1994 OJTE -- 1989- Supervisor with training as collateral duty -- 1994-present I also maintain an "FPL development" binder to assist controllers in career and personal development after FPL certification.
4	Reached full performance level in 1980 after 5 years of work/training. Served as OJT instructor from 1981 until 1986. Spent 18 months in traffic management, 1988 until 1992, controller and OJT instructor. 1992 until present -- Supervisor performing controller evaluations and certifications.
5	Crew training specialist. Various details to the training department for upgrade training classes. Evaluator and certifier.
6	OJT instructor -- 1974-1979, 1982-1985 Area Supervisor with collateral duty as Area Training Supervisor -- 1985-1996
7	Controller evaluations/every 6 months/for 10 years. Skill checks and evaluations on ATC developmentals -- 15 years/skill check -- certifications on going.
8	Hired in the FAA, trained in a nonradar environment (later radar), became a Full Performance ATCS, and worked in Traffic Management prior to becoming a supervisor in the same facility. I am a pilot with 1000+ hours of flight time.
9	1 year training developmentals in FSS option 4 years training developmentals in En Route option 10 years evaluating developmentals in En Route option -- some/part of above involved in manual and/or Dysim lab -- Area Training Supervisor -- 2 years

Appendix E

Participation Consent Form

January 28, 1997

I, _____, agree to participate in the Video Tape

(print full name)

Performance Rating Project, which is being conducted January 28-February 7, 1997, at the Federal Aviation Administration William J. Hughes Technical Center. I agree that portions of this activity may be audio-taped or video-taped. I understand that my contribution will be held confidential and used for research purposes only.

_____

Signature

Appendix F

Final Questionnaire

Observer Code _____ Date _____

A. Indicate the importance of the 6 performance areas to overall air traffic control performance by selecting a weight score (between 0 and 100) for each area. Higher weights indicate more important performance areas. Your overall performance rating for each area will be multiplied by your indicated weight to compute a weighted overall performance score for each scenario. The weights must sum to 100.

EXAMPLE:

20	MAINTAINING SAFE AND EFFICIENT TRAFFIC FLOW
20	MAINTAINING ATTENTION AND SITUATION AWARENESS
20	PRIORITIZING
20	PROVIDING CONTROL INFORMATION
10	TECHNICAL KNOWLEDGE
10	COMMUNICATING
100	

YOUR SELECTIONS:

_____	MAINTAINING SAFE AND EFFICIENT TRAFFIC FLOW
_____	MAINTAINING ATTENTION AND SITUATION AWARENESS
_____	PRIORITIZING
_____	PROVIDING CONTROL INFORMATION
_____	TECHNICAL KNOWLEDGE
_____	COMMUNICATING
100	

FINAL QUESTIONNAIRE
(continued)

Videotape evaluations of controllers is a new methodology that has not been done in previous research. In order to evaluate and improve this methodology, we would like your opinions regarding the following questions.

1. As compared to viewing controllers "live", the radar display showed sufficient information for me to make my evaluations.

1	2	3	4	5	6	7	8	9	10
strongly disagree									strongly agree

2. The training period was sufficient for me to become familiar with the new evaluation form.

1	2	3	4	5	6	7	8	9	10
strongly disagree									strongly agree

3. Please write down any recommendations you have for improving the video tape evaluations methodology (e.g., training format, video tape presentation, etc.).

4. Please list any other objective performance measures that should be collected to evaluate controller effectiveness (e.g., aircraft flight time, aircraft fuel consumption).

5. Please discuss which aspects of the controller performance video evaluation study could be improved in future efforts.
   a. Pseudo-pilot video tape (presented on left screen)
   b. ATCoach replay (presented on center screen)
   c. Over-the-shoulder video tape (presented on right screen)
   d. Audio
   e. Other

6. Please rate the professionalism and personal demeanor of the lab personnel involved in conducting this study.

7. How can R&D help operations at your facility?

Appendix G

Hourly Schedule of Activities

Hourly Schedule of Activities for Training Session I (Day #2)

Time Block	Scheduled Activity
8:00 - 9:30	Orientation (EARL and PAUL)
	Introductions, Background Questionnaire, and Lab Tour
9:30 - 9:45	15-Minute Break
9:45 - 10:45	Performance Measurement, Project Overview, and Previous Experiments (EARL and PAUL)
10:45 - 11:00	15-Minute Break
11:00 - 12:00	16 Personality Factor (JENNY)
12:00 - 1:00	Lunch Break
1:00 - 2:30	ZJX Sector (DAVE)
	Layout and Procedures, Review of Synopsis, and Hands-On Demo Scenario
2:30 - 2:45	15-Minute Break
2:45 - 3:45	Discussion of Factors and Issues Most Critical to En Route Air Traffic Control (LAURIE)
3:45 - 4:00	Summary and Question Period (EARL, PAUL, JENNY, DAVE, and LAURIE)
	Done for the Day

Hourly Schedule of Activities for Training Session II, Day 1 (Day #3)

Time Block	Scheduled Activity
8:00 - 9:00	Overview of Evaluation Form (DAVE and JENNY)
	Purpose, Background Work, Performance Scales, and Design Features
9:00 - 9:15	15-Minute Break
9:15 - 10:15	Presentation of Practice Replay (DAVE)
	View “A” Scenario Segment, Review Sector and Procedures, and Use Evaluation Form
10:15 - 10:30	15-Minute Break
10:30 - 11:30	Group Discussion of Evaluations (DAVE and JENNY)
	Discuss Performance Areas, Criteria Standards, and Scenario Ratings
11:30 - 12:30	Lunch Break
12:30 - 1:30	Presentation of Practice Replay (DAVE)
	View “B” Scenario Segment, Review Sector and Procedures, and Use Evaluation Form
1:30 - 1:45	15-Minute Break
1:45 - 2:45	Group Discussion of Evaluations (DAVE and JENNY)
	Discuss Performance Areas, Criteria Standards, and Scenario Ratings
2:45 - 3:00	15-Minute Break
3:00 - 4:00	Summary and Question Period (DAVE and JENNY)
	Done for the Day

Hourly Schedule of Activities for Training Session II, Day 2 (Day #4)

Time Block	Scheduled Activity
8:00 - 9:00	Presentation of Practice Replay (DAVE)
	View "B" Scenario Segment, Review Sector and Procedures, and Use Evaluation Form
9:00 - 9:15	15-Minute Break
9:15 - 10:15	Group Discussion of Evaluations (DAVE and JENNY)
	Discuss Performance Areas, Criteria Standards, and Scenario Ratings
10:15 - 10:30	15-Minute Break
10:30 - 12:00	Discussion of Generic En Route Sector (STAN and DAVE)
	Layout and Procedures
12:00 - 1:00	Lunch Break
1:00 - 2:00	Presentation of Practice Replay (DAVE)
	View "A" Scenario Segment, Review Sector and Procedures, and Use Evaluation Form
2:00 - 2:15	15-Minute Break
2:15 - 3:15	Group Discussion of Evaluations (DAVE and JENNY)
	Discuss Performance Areas, Criteria Standards, and Scenario Ratings
3:15 - 3:30	15-Minute Break
3:30 - 4:00	Summary and Question Period (DAVE and JENNY)
	Done for the Day

Hourly Schedule of Activities for Training Session II, Day 3 (Day #5)

Time Block	Scheduled Activity
8:00 - 9:00	Presentation of Practice Replay (DAVE)
	View "A" Scenario Segment, Review Sector and Procedures, and Use Evaluation Form
9:00 - 9:15	15-Minute Break
9:15 - 10:15	Group Discussion of Evaluations (DAVE and JENNY)
	Discuss Performance Areas, Criteria Standards, and Scenario Ratings
10:15 - 10:30	15-Minute Break
10:30 - 11:30	Discussion of Generic En Route Sector (STAN and DAVE)
11:30 - 1:00	Lunch Break
1:00 - 2:00	Presentation of Practice Replay (DAVE)
	View "B" Scenario Segment, Review Sector and Procedures, and Use Evaluation Form
2:00 - 2:15	15-Minute Break
2:15 - 3:15	Group Discussion of Evaluations (DAVE and JENNY)
	Discuss Performance Areas, Criteria Standards, and Scenario Ratings
3:15 - 3:30	15-Minute Break
3:30 - 4:00	Summary and Question Period (DAVE, JENNY, and STAN)
	Done for the Day

Hourly Schedule of Activities for an Evaluation Day (Days #6-#8)

Time Block	Scheduled Activity
8:00 - 8:30	Getting Settled / Question Period
8:30 - 9:30	View Replay
9:30 - 9:50	Finish Evaluation Form
9:50 - 10:05	15-Minute Break
10:05 - 11:05	View Replay
11:05 - 11:25	Finish Evaluation Form
11:25 - 12:30	Lunch Break
12:30 - 1:30	View Replay
1:30 - 1:50	Finish Evaluation Form
1:50 - 2:05	15-Minute Break
2:05 - 3:05	View Replay
3:05 - 3:25	Finish Evaluation Form
3:25 - 3:40	15-Minute Break
3:40 - 4:00	Discussion of Ratings and Scenarios
	Done for the Day

Hourly Schedule of Activities for the Evaluation Day with Debriefing (Day #9)

Time Block	Scheduled Activity
8:00 - 8:20	Getting Settled / Question Period
8:20 - 8:40	Final Questionnaire
8:40 - 8:55	15-Minute Break
8:55 - 9:55	View Replay
9:55 - 10:15	Finish Evaluation Form
10:15 - 10:30	15-Minute Break
10:30 - 11:30	View Replay
11:30 - 11:50	Finish Evaluation Form
11:50 - 12:50	Lunch Break
12:50 - 1:50	View Replay
1:50 - 2:10	Finish Evaluation Form
2:10 - 2:25	15-Minute Break
2:25 - 3:25	View Replay
3:25 - 3:45	Finish Evaluation Form
3:45 - 4:00	Discussion of Ratings and Scenarios
	Debriefing
	Done for the Day

Appendix H

Summary Sheet

I - MAINTAINING SAFE AND EFFICIENT TRAFFIC FLOW

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

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

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

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

II - MAINTAINING ATTENTION AND SITUATION AWARENESS

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

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

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

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

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

III - PRIORITIZING

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

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

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

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

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

IV - PROVIDING CONTROL INFORMATION

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

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

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

V - TECHNICAL KNOWLEDGE

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

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

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

VI – COMMUNICATING

21. Using Proper Phraseology..... **1 2 3 4 5 6 7 8**
    - using words and phrases specified in the 7110.65
    - using phraseology that is appropriate for the situation
    - using minimum necessary verbiage
    - speaking with confident, authoritative tone of voice

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

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

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

Appendix I

Presentation Order of Scenarios

Day 1		Day 2		Day 3		Day 4	
Controller	Scenario (condition)	Controller	Scenario (condition)	Controller	Scenario (condition)	Controller	Scenario (condition)
1	1(A)*	5	5(B)	5	8(A)	4	4(B)[a]
2	2(B)**	7	6(A)	8	9(B)	3	3(A) [a]
3	3(A)	9	7(B)	6	1(A)	8	9(B) [a]
4	4(B)	10	1(A)	1	10(B)	7	6(A) [a]

* Condition A – Current ATC procedures

** Condition B – Current ATC procedures but direct routings included

[a] Repeated scenarios

Appendix J

System Effectiveness Measures

Abbreviation	Description
NCNF	Number of Conflicts (less than 5 nm and 2,000 feet separation)
NALT	Number of Altitude Assignments
NHDG	Number of Heading Assignments
NSPD	Number of Speed Assignments
NPTT	Number of Ground-to-Air Transmissions
CMAV	Cumulative Average of System Activity/Aircraft Density (number of aircraft within 8 nm of another aircraft)
ATWIT	Air Traffic Workload Input Technique Rating

Appendix K

16PF Descriptive Statistics

Table K-1. Descriptive Statistics for Participant Scores on 16PF Global Factors

Factor	Mean	Standard Deviation
Extraversion	4.33	2.78
Anxiety	6.11	2.26
Tough-Mindedness	6.44	2.13
Independence	5.00	1.94
Self Control	5.44	2.07

Table K-2. Descriptive Statistics for Participant Scores on 16PF Basic Factors

Factor	Mean	Standard Deviation
Warmth	3.56	1.88
Reasoning	7.44	2.07
Emotional Stability	5.78	1.79
Dominance	5.11	1.96
Liveliness	5.33	2.00
Rule Conscious	5.78	1.72
Social Boldness	4.56	2.01
Sensitivity	4.44	1.51
Vigilance	5.56	1.13
Abstractness	5.56	2.07
Privateness	5.22	2.33
Apprehension	6.33	1.94
Openness to Change	5.33	1.87
Self Reliance	7.00	2.45
Perfectionism	4.56	2.13
Tension	6.11	2.42

Appendix L
Correlational Analysis Between Participant Ratings and Scores
on 16PF Global Factors

RATING SCALES	16PF GLOBAL FACTORS: Extroversion	Anxiety	Tough-Mindedness	Independence	Self Control
1. Maintaining Separation and Resolving Potential Conflicts	.17	-.22	-.07	.13	-.16
2. Sequencing Arrival, Departure, and En Route Aircraft Efficiently	.48*	-.24*	-.38*	.26*	-.49*
3. Using Control Instructions Effectively/Efficiently	.09	-.10	-.13	.12	-.21
4. Overall Safe and Efficient Traffic Flow Scale Rating	.30*	-.25*	-.17	.17	-.30*
5. Maintaining Situational Awareness	.32*	-.22	-.14	.27*	-.30*
6. Ensuring Positive Control	.41*	-.29*	-.32*	.31*	-.40*
7. Detecting Pilot Deviations from Control Instructions	.17	-.21	-.12	.14	-.18
8. Correcting Errors in a Timely Manner	.15	-.14	.04	-.04	-.16
9. Overall Attention and Situation Awareness Scale Rating	.39*	-.31*	-.23	.31*	-.38*
10. Taking Actions in an Appropriate Order of Importance	.08	-.04	-.22	.11	-.15
11. Preplanning Control Actions	.22	-.06	-.44*	.27*	-.30*
12. Handling Control Tasks for Several Aircraft	.06	-.09	-.19	18	-.16
14. Overall Prioritizing Scale Rating	.20	-.14	-.33*	.27*	-.27*
15A. Providing Essential Air Traffic Control Information	.44*	-.21	-.34*	.20	-.46*
15B. Providing Additional Air Traffic Control Information	.44*	-.21	-.25*	.16	-.36*
16. Providing Coordination	.18	-.06	-.21	.02	-.25*
12. Overall Providing Control Information Scale Rating	.43*	-.23	-.28*	.14	-.43*
19A. Showing Knowledge of Aircraft Capabilities and Limitations	.33	-.15	-.30	.22	-.49*
19B. Showing Effective Use of Equipment	.29	-.16	-.11	.07	-.31*
20. Overall Technical Knowledge Scale Rating	.31*	-.12	-.27	.16	-.41*
21. Using Proper Phraseology	.24*	-.16	-.22	.17	-.29*
22. Communicating Clearly and Efficiently	.25*	-.22	-.23	.24*	-.31*
23. Listening to Pilot Readbacks and Requests	.33*	-.24*	-.29*	.36*	-.30*
24. Overall Communicating Scale Rating	.32*	-.24*	-.25*	.26*	-.36*
Overall Weighted Performance Score	.33*	-.29*	-.19	.22	-.32*

* = $p < .05$

Note: Items 13 and 18 are not shown here because participants did not rate controllers on them.

Appendix M

Correlational Analysis Between Participant Ratings and Scores on 16PF Basic Factors

RATING SCALES	16PF BASIC FACTORS					
	Reasoning	Rule Conscious	Vigilance	Openness to Change	Perfectionism	Tension
1. Maintaining Separation and Resolving Potential Conflicts	-.06	-.18	-.09	.08	-.13	-.19
2. Sequencing Arrival, Departure, and En Route Aircraft Efficiently	.17	-.45*	-.12	.34*	-.42*	-.22
3. Using Control Instructions Effectively/ Efficiently	-.02	-.14	.05	.11	-.24*	-.11
4. Overall Safe and Efficient Traffic Flow Scale Rating	.01	-.29*	-.13	.19	-.26*	-.20
5. Maintaining Situational Awareness	-.13	-.22	.03	.31*	-.38*	-.18
6. Ensuring Positive Control	.15	-.42*	-.19	.30*	-.31*	-.20
7. Detecting Pilot Deviations from Control Instructions	.00	-.20	-.07	.06	-.14	-.19
8. Correcting Errors in a Timely Manner	.00	-.11	-.23	-.09	-.20	.01
9. Overall Attention and Situation Awareness Scale Rating	.02	-.35*	-.13	.33*	-.40*	-.22
10. Taking Actions in an Appropriate Order of Importance	.23	-.20	-.09	.08	-.04	-.02
11. Preplanning Control Actions	.36*	-.34*	-.02	.22	-.13	-.08
12. Handling Control Tasks for Several Aircraft	.13	-.17	-.03	.09	-.10	-.04
14. Overall Prioritizing Scale Rating	.22	-.29*	-.04	.26*	-.18	-.12
15A. Providing Essential Air Traffic Control Information	.15	-.42*	-.12	.25*	-.39*	-.19
15B. Providing Additional Air Traffic Control Information	.11	-.37*	-.18	.08	-.23	-.13
16. Providing Coordination	.14	-.26*	-.09	.02	-.12	-.07
17. Overall Providing Control Information Scale Rating	.12	-.42*	-.19	.17	-.31*	-.19
19A. Showing Knowledge of Aircraft Capabilities and Limitations	-.07	-.40*	.01	.25	-.39*	-.09
19B. Showing Effective Use of Equipment	-.07	-.25	-.12	.01	-.28	-.10
20. Overall Technical Knowledge Scale Rating	.02	-.34*	-.04	.09	-.30*	-.08
21. Using Proper Phraseology	.17	-.28*	-.16	.15	-.25*	-.08
22. Communicate Clearly and Efficiently	-.01	-.30*	-.02	.23	-.26*	-.21
23. Listening to Pilot Readbacks and Requests	.07	-.29*	.02	.43*	-.31*	-.24*
24. Overall Communicat. Scale Rating	.04	-.33*	-.09	.26*	-.33*	-.19
Overall Weighted Performance Score	.03	-.33*	-.16	.21	-.28*	-.22

* = $p < .05$

Note: Items 13 and 18 are not shown here because participants did not rate controllers on them.