otе technical note tech

Display System Replacement Baseline Research Report

Kenneth R. Allendoerfer, ACT-530
Joseph Galushka, ACT-530
Richard H. Mogford, Ph.D., ACT-530

December 2000
DOT/FAA/CT-TN00/31

DISTRIBUTION STATEMENT A
Approved for Public Release
Distribution Unlimited

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

U.S. Department of Transportation
Federal Aviation Administration
William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

NOTICE

Technical Report Documentation Page

1. Report No. DOT/FAA/CT-TN00/31	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Display System Replacement Baseline Research Report		**5. Report Date** December 2000
		6. Performing Organization Code ACT-530
7. Author(s) Kenneth R. Allendoerfer, Joseph Galushka, and Richard H. Mogford, Ph.D., ACT-530		**8. Performing Organization Report No.** DOT/FAA/CT-TN00/31
9. Performing Organization Name and Address Federal Aviation Administration William J. Hughes Technical Center Atlantic City International Airport, NJ 08405		**10. Work Unit No. (TRAIS)**
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address Federal Aviation Administration Human Factors Division 800 Independence Ave., S.W. Washington, DC 20591		**13. Type of Report and Period Covered** Technical Note
		14. Sponsoring Agency Code AAR-100
15. Supplementary Notes		

16. Abstract

This report provides baseline measurements on the Display System Replacement (DSR). These measurements followed six constructs: safety, capacity, performance, workload, usability, and simulation fidelity. To collect these measurements, human factors researchers conducted an air traffic control simulation using four sectors of Washington Air Route Traffic Control Center (ARTCC) airspace with a traffic volume representing a 90th percentile day. Ten controllers from Washington ARTCC served as participants in the study. Recordings and questionnaires provided objective and subjective measurements such as the number of aircraft controlled and controller workload. This report provides statistics at several levels of specificity: aggregated across all sectors and positions, by individual sectors and positions, and by 12-minute intervals. We intend that data from the study will provide a meaningful representation of the DSR controller position. In addition, this report compares these data to equivalent data collected on the Plan View Display, the equipment that the DSR replaces. We discuss several differences between the systems and provide recommendations for improvements to the baseline process.

17. Key Words En Route Air Traffic Control Display System Replacement Performance Baseline		**18. Distribution Statement** This document is available to the public through the National Technical Information Service, Springfield, Virginia, 22161	
19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 110	**22. Price**

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

ACKNOWLEDGMENTS

The simulations discussed here required the efforts of many people and organizations. Though not listed by name, we would first like to thank our controller participants and subject matter expert observers from Washington Air Route Traffic Control Center (ZDC). Without their dedicated efforts, these simulations could not have been conducted. We would also like to thank the management of ZDC and the National Air Traffic Controllers Association for coordinating the participation of the field personnel.

In addition, the following individuals provided valuable assistance during the planning and conduct of the Display System Replacement Baseline, the data reduction and analysis, or the preparation of this report.

Federal Aviation Administration

Clark Desing, Salt Lake City ARTCC
John Frederick, ACT-231
Rex Jackson, ZDC
Doug Lance, ZDC
Albert Macias, ACT-510
Robert Oliver, ACT-510
Jeffrey Phillips, Indianapolis ARTCC

Computer Sciences Corporation

Michael Dunleavy

Harris Corporation

Leo Smith
Doug Gerdes

Federal Data Corporation

Jean Dunn

Raytheon Corporation

Scott Cramer
Michael Ross

System Resources Corporation

Deborah Hough
George Rowand
Otto Smith

Joseph Galushka is now at HMI Solutions, Smithville, NJ. Richard Mogford is now at the NASA Ames Research Center, Moffett Field, CA.

Table of Contents

List of Illustrations

EXECUTIVE SUMMARY

The Federal Aviation Administration is in the midst of deploying the Display System Replacement (DSR) to Air Route Traffic Control Centers (ARTCCs) nationwide. In support of this effort, the William J. Hughes Technical Center National Airspace System Human Factors Branch conducted a baseline simulation of en route air traffic control operations using the DSR. The simulation provided data for five operational constructs: Safety, Capacity, Performance, Workload, and Usability and for a sixth, non-operational construct, Simulation Fidelity. These constructs are the same as those used in the Plan View Display (PVD) Baseline conducted in 1995.

The DSR Baseline also used the same airspace, traffic scenarios, and controller participants as the PVD Baseline. The simulation used four Washington ARTCC sectors and two traffic scenarios that represented a 90th percentile day for traffic volume. Six controllers who participated in the PVD Baseline also participated in the current study. Some differences in methodology between the baselines included different simulation platforms, different communication equipment, and different pseudopilots.

The DSR Baseline also used the same data collection and analysis techniques as the PVD Baseline. Human factors researchers collected objective data from the output of the simulation platform and the communication system. We collected subjective data using controller and expert observer questionnaires. We measured subjective controller workload using the Air Traffic Workload Input Technique. We reduced the data using the same methods as the PVD Baseline whenever possible. We report the data here at the overall, individual sector, and 12-minute interval levels.

In addition to the DSR Baseline data, this report presents a comparison of the DSR and PVD Baselines. A seven-member Operational Review Team examined the data from both baselines to ensure validity and usefulness. The team developed rationales for any differences found between the baselines and conducted further analyses when needed. Some important differences were more data block positioning, halo initiations, data entries, and data entry errors in the DSR Baseline and higher workload ratings in the PVD Baseline. The review team also eliminated some data based on validity concerns and made recommendations for improving the baseline process. Some reasons for elimination were differences in the simulation platform, differences in the procedures, and differences in the data reduction and analysis.

The review team generated recommendations for the DSR program. The team recommended further research and possible improvements to DSR data block readability, flight strip bays, keyboards, and vector-line controls. They also generated recommendations for improving the baseline process including increasing configuration management, using side-by-side comparisons of systems, and using scenarios that are more complex.

1. Introduction

In January 1999, the Federal Aviation Administration (FAA) formally dedicated the Display System Replacement (DSR) at Seattle Air Route Traffic Control Center (ARTCC). The DSR program is part of a larger effort to modernize the FAA en route Air Traffic Control (ATC) system and will be operational in all ARTCCs by early 2000. The DSR replaces the Plan View Display (PVD) that had been used for data display and entry since the 1970s. The DSR will improve system reliability and maintainability and will provide capacity for future en route enhancements such as improved weather information and conflict probe (FAA, 1996).

1.1 Background

As part of its test and evaluation activities, the FAA sponsored human performance baseline simulations for the PVD and the DSR. The PVD Baseline was completed in 1995, and the results are reported in the *Plan View Display Baseline Research Report* (Galushka, Frederick, Mogford, & Krois, 1995). For this original study, a suite of metrics that quantified the operational efficiency and effectiveness of en route ATC systems was developed. Those baseline metrics were applied to the PVD in a realistic human-in-the loop simulation at the William J. Hughes Technical Center using Washington ARTCC (ZDC) airspace and traffic scenarios. In 1997, human factors researchers from the National Airspace System Human Factors Branch (ACT-530) applied the same baseline metrics to the DSR during another human-in-the-loop simulation. We followed the PVD Baseline methodology as closely as possible and used the same airspace, traffic scenarios, data collection and analysis techniques, and many of the same controller participants. The data collected in the original study form one half of the comparison reported here.

1.2 Purpose

The purpose of this report is fourfold:

a. It presents data collected during the DSR Baseline using the suite of metrics developed for the PVD Baseline. We collected these data during human-in-the-loop simulations of en route operations with controllers using the DSR. These data quantify the operational efficiency and effectiveness of the DSR.

b. It presents a comparison of the DSR Baseline data to data collected during the PVD Baseline. The researchers and subject matter experts (SMEs) reviewed the comparison for validity and usefulness. This represents the first comparison of objective and subjective data collected for the PVD and DSR under equivalent, realistic simulation conditions.

c. It presents recommendations for the DSR program about research that should be conducted into particular aspects of the DSR. The recommended research may lead to future DSR upgrades and improvements.

d. It presents recommendations for how the baseline process can be improved.

2. Operational Constructs and Baseline Metrics

In 1995, Air Traffic Requirements (now part of the Air Traffic System Requirements Service [ARS]) identified five high-level operational constructs upon which to base evaluations of the efficiency and effectiveness of ATC systems: Safety, Capacity, Performance, Workload, and Usability. For the PVD Baseline, engineering research psychologists, en route Air Traffic Control Specialists (ATCSs), and other ATC automation, training, and management SMEs added a sixth construct, Simulation Fidelity, to measure the realism and accuracy of baseline simulations. For each of the constructs, they developed several baseline metrics for which objective and subjective data could be obtained. For more information about this process, see the *Plan View Display Baseline Research Report* (Galushka et al., 1995).

In the current study, we collected data for the DSR following the original constructs. We defined the constructs as follows:

a. Safety represents the extent to which the system allows aircraft to traverse a section of airspace without a dangerous incident such as a violation of applicable separation minima.

b. Capacity represents the amount of traffic that the system allows to safely and efficiently traverse a section of airspace during a period of time.

c. Performance represents the amount and quality of user interaction with the system.

d. Workload represents the cognitive and physical task demands of the system as experienced by its users.

e. Usability represents how easily particular aspects of the system such as controls and displays can be learned and used for their intended purpose.

f. Simulation Fidelity represents characteristics of the traffic scenarios and laboratory environment and simulation participant opinions about the realism and accuracy of the simulation.

As part of the preparations for the DSR Baseline, we re-evaluated each metric to determine its applicability to the DSR. When appropriate, we modified the data collection technique or eliminated the metric. Complete descriptions of these metrics can be found in the *Air Traffic Control System Baseline Methodology Guide* (Allendoerfer & Galushka, 1999). We collected data for the following metrics:

a. Safety

1. Operational Errors. This measure represents the total number of violations of applicable separation minima.

2. Conflict Alerts. This measure represents the total number of warnings issued to controllers about imminent separation violations. These warnings are issued by the Host Computer System (HCS) according to FAA algorithms.

3. Halo Initiations. This measure represents the total number of times a controller initiated the display of the halo (also known as the J-Ring).

4. Data Block Positioning. This measure represents the total number of times a controller changed leader-line lengths and leader-line directions to maintain data block readability.

5. Other Safety-Critical Issues. This measure represents SME observations of safety-related issues and deficiencies.

b. Capacity

1. Aircraft Under Control. This measure represents the total number of aircraft receiving ATC services from a controller.

2. Time in Sector. This measure represents the average time aircraft spend in a particular sector.

c. Performance

1. Overall Data Entries. This measure represents the number of data entries made by a controller using the keyboard and/or trackball across all data entry types.

2. Specific Data Entry Types. This measure represents the number of data entries made by a controller using the keyboard and trackball for specific data entry types.

3. Data Entry Errors. This measure represents the total number of data entry error messages returned by the HCS.

4. Number of Altitude, Speed, and Heading Changes. This measure represents the total number of controller-initiated altitude, speed, and heading changes made by simulated aircraft.

5. Self-Assessments of Performance. This measure represents subjective performance ratings given by a controller participant at the end of a simulation run. Ratings range from 1 (low) to 8 (high). The measure comprises two submeasures:

   a) Quality of ATC services from a controller point of view

   b) Quality of ATC services from a pilot point of view

6. Observer Assessments of Performance. This measure represents ratings of participant performance during a simulation run made by one or more SME observers. Ratings range from 1 (Least Effective) to 8 (Most Effective). The measure comprises six submeasures with three to five rating scales each. In past baselines, we have reported data for only the overall items for each submeasure. These items are as follows:

   a) Maintaining Safe and Efficient Traffic Flow

   b) Maintaining Attention and Situation Awareness

   c) Prioritizing

   d) Providing Control Information

   e) Technical Knowledge

   f) Communicating

d. Workload

1. Air Traffic Workload Input Technique (ATWIT) Workload. This measure represents the subjective workload ratings given by the participants during a specific time interval. Ratings range from 1 (low) to 7 (high).

2. Post-Run Workload. This measure represents subjective workload ratings given by the controller participants at the end of the simulation run. Ratings range from 1 (low) to 8 (high).

3. Communication Taskload. This measure represents the total number of controller-initiated push-to-talk (PTT) air-ground communications (i.e., communications between a controller and the pseudopilots working traffic in his or her sector).

4. Coordination Taskload. This measure represents the total number of controller-initiated PTT ground-ground communications (i.e., communications between a controller and controllers working in other sectors or ghost sectors).

e. Usability

We based this construct on controller responses on the Final Questionnaire, Section A. Ratings range from 1 (low) to 8 (high). The construct includes the following questionnaire items:

1. Flight Progress Strip Access
2. Flight Progress Strip Read/Mark
3. Ease of Access of Controls
4. Operation of Controls Intuitive
5. Keyboard Ease of Use
6. Radar and Map Displays Ease of Reading
7. Radar and Map Displays Ease of Understanding
8. Workstation Space
9. Equipment, Displays, and Controls Support Efficient ATC
10. Equipment, Displays, and Controls Impose Limitations
11. Equipment, Displays, and Controls Overall Effectiveness

f. Simulation Fidelity

1. Traffic Scenario Characteristics. This measure represents important features of the traffic scenarios used in the simulation and consists of several submeasures. They are

   a) length of each scenario,

   b) total number of arrivals,

   c) total number of departures,

   d) total number of overflights,

e) total number of propeller aircraft, and

f) total number of jet aircraft.

2. Realism Rating. This measure represents the perceived realism and fidelity of the simulation run as rated by a controller participant. Ratings range from 1 (Not Very Realistic) to 7 (Extremely Realistic).

3. Impact of Technical Problems Rating. This measure represents the perceived impact of technical problems on the participants' ability to control traffic during the simulation run. Ratings range from 1 (Not Very Much) to 8 (A Great Deal).

4. Impact of Pseudopilots Rating. This measure represents the perceived impact of the pseudopilots on the participants' ability to control traffic during the simulation run. Ratings range from 1 (Not Very Much) to 8 (A Great Deal).

5. Scenario Difficulty Rating. This measure represents the perceived difficulty of the traffic scenario as rated by participants. Ratings range from 1 (Not Very Difficult) to 8 (Extremely Difficult).

3. Method

To make valid comparisons, data must be collected under equivalent conditions and analyzed using equivalent methods. For the DSR Baseline, we adapted the PVD Baseline methodology to the DSR platform and to the improved simulation capabilities at the Technical Center. We used the same airspace, traffic scenarios, and many of the same controllers and observers. To the extent possible, we used the same data collection instruments, analysis tools, and reporting style. In the following sections, we describe the DSR Baseline methodology and note any differences from the PVD Baseline. The general baseline methodology can be found in the *Air Traffic Control System Baseline Methodology Guide* (Allendoerfer & Galushka, 1999).

3.1 Personnel

3.1.1 ATCS Participants

Ten Full Performance Level (FPL) ATCSs from ZDC served as participants. Six had served previously as participants in the PVD Baseline. All participants were current and certified on the ZDC sectors used in the baseline. At the time of the baseline, the participants had already completed a DSR training course at the FAA Display Development Facility and had just completed 2 weeks of the DSR Operational Test and Evaluation (OT&E). Table 1 presents demographic information about the participants collected on the Background Questionnaire (Appendix A).

Table 1. Background Questionnaire Data

Questionnaire Item	Result (where applicable, ratings are on 1-8 scale)
Age	$M = 34.6$, $SD = 4.14$
Years experience controlling traffic	$M = 12.3$, $SD = 5.10$
Months in the last year actively controlling traffic	$M = 12.0$, $SD = 0.00$
Hours experience with the DSR	$M = 22.8$, $SD = 12.80$
Current position	All FPL
Domain with most experience	All en route
Corrective lenses	Five wore corrective lenses
Current state of health	$M = 7.3$, $SD = 1.06$
Current skill as an ATCS	$M = 7.4$, $SD = 0.70$
Level of experience with personal computers	$M = 5.6$, $SD = 1.58$
Level of satisfaction with the DSR	$M = 3.6$, $SD = 1.17$

3.1.2 Subject Matter Expert Observers

Two supervisory-level ATCSs from ZDC served as SME Observers in the DSR Baseline. One had served previously as an SME Observer in the PVD Baseline.

3.2 Facilities, Equipment, and Materials

3.2.1 Display System Replacement Laboratory

We conducted the DSR Baseline in the DSR Laboratory at the Technical Center, Building 316. The controllers staffed two sectors consisting of one radar (R) and one data (D) position each. Assistant (A) positions for each sector were also available but were not staffed. Each R position included a Sony 20-inch by 20-inch Main Display Monitor, an R-position keyboard, a three-button trackball, and two Voice Switching Control System (VSCS) panels. Each D position included a 15-inch color monitor showing the D position Computer Readout Display (CRD), a D-position keyboard, two VSCS panels, and several flight strip bays. Flight strip bays on the A positions were also available for use.

The DSR Baseline used the VSCS rather than the Amecom system for air-ground and ground-ground communications. At the time of the PVD Baseline, the VSCS had not been deployed to the field. However, in the interim between baselines, the VSCS was deployed, and all the participants had extensive training and experience with it by the time of the DSR Baseline. In addition, the DSR was engineered to operate in conjunction only with the VSCS and not with the legacy voice switch systems. Though this difference in voice switch equipment does make it more difficult to compare the PVD and DSR Baselines, we believe that using the VSCS in the DSR Baseline was necessary to preserve realism and external validity.

One ghost sector was located behind the operational sectors. Simulation support personnel staffed the ghost sector, which was responsible for handoffs and coordination with the simulated sectors. The ghost sector played the role of all sectors and facilities not staffed by the participants.

3.2.2 Target Generation Facility

The DSR Baseline used the Target Generation Facility (TGF) for scenario generation. The TGF provided a realistic simulation of ZDC traffic including complex aircraft and pilot behavior. Professional pseudopilots played the role of pilots in the scenario. The pseudopilots communicated with the controllers via the VSCS and issued commands to the simulated aircraft when cleared by the controllers.

The DSR Baseline used the TGF rather than the HCS Dynamic Simulation (DYSIM) capability for scenario generation. In the PVD Baseline, ZDC controllers filled the pseudopilot and ghost sector roles when not serving as study participants. That technique can be beneficial in that controllers are knowledgeable about aircraft behavior and can provide realistic, adaptive communications. However, that technique also can reduce the repeatability of simulations because controller-pseudopilots sometimes may alter aircraft routes and behavior at their discretion. Professional pseudopilots will not take such discretionary actions unless required by the simulation methodology. Because the TGF provided superior scenario realism and because all other DSR OT&E activities used the TGF, we used the TGF in the DSR Baseline.

3.2.3 Washington ARTCC Airspace

The DSR Baseline simulated the ZDC sectors used in the PVD Baseline. Descriptions of the sectors at ZDC are listed below, and any differences between the actual and simulated sectors are noted.

a. Sector 26, known as Sampson, is a low-altitude sector responsible for altitudes between 11,000 ft to 23,000 ft. Sampson borders Jacksonville ARTCC and is completely bordered beneath by terminal airspace. Controllers staffing Sampson interface with the following approach control facilities: Fayetteville, Raleigh-Durham, Seymour Johnson, Wilmington, and Patuxent River. A large portion of the traffic in this sector are Raleigh-Durham International Airport (RDU) southbound departures.

b. Sector 27, known as Liberty, is a low-altitude sector responsible for altitudes 11,000 ft to 23,000 ft. Liberty borders Atlanta ARTCC and interfaces with Greensboro, Raleigh-Durham, and Fayetteville approach control facilities. This sector handles numerous traffic flows including RDU westbound and northbound departures, RDU arrivals from the southwest and south, Charlotte/Douglas International Airport (CLT) northbound and eastbound departures, and CLT arrivals from the east. Liberty also handles military traffic from Pope Air Force Base.

c. Sector 35, known as Wilmington, was combined with sector 09, known as Dixon, during the DSR Baseline, as is often done in the field. This combined high/ultra-high altitude sector is responsible for altitudes 24,000 ft and above. This combined sector handles primarily northbound and southbound traffic from airports in Florida, New York, New England, and Pennsylvania.

d. Sector 38, known as Tar River, is a high-altitude sector responsible for altitudes 24,000 ft and above. Tar River handles primarily northbound traffic, particularly arrival flows for the three major airports in the Washington-Baltimore area. This sector also transitions RDU departures to the south and east from the Rocky Mount and Sampson sectors to high altitude strata. Other major traffic flows are from New York, New England, and Pennsylvania airports southbound.

3.2.4 Traffic Scenarios

The DSR Baseline used two traffic scenarios based on the scenarios used in the PVD Baseline. The first scenario used adjacent sectors 26 (low) and 38 (high) and contained 70 min of traffic. The second scenario used non-adjacent sectors 27 (low) and 35 (high) and contained 100 min of traffic. In both scenarios, the first 10 min were excluded from the data to allow the traffic volume to increase to a realistic level.

The original traffic scenarios were developed for DYSIM using System Analysis Recording (SAR) flight data recorded at ZDC in September 1992. The scenarios were recorded on a 90th percentile day for traffic volume, which we believed at that time to be sufficient to functionally exercise the PVD. These scenarios were verified and rated by a ZDC SME and tested in the Technical Center laboratories. Unusual events such as emergencies or operational errors were purposely removed from the scenarios to preserve repeatability of the scenarios and to focus the baselines on routine ATC operations rather than on techniques for handling problems.

Prior to the DSR OT&E, TGF personnel adapted the DYSIM scenarios to run on the TGF simulation platform. This required some minor modifications to the scenarios, primarily to improve simulator performance and to eliminate inconsistencies. We believe that none of these modifications had any impact on the traffic seen by controllers during the simulation runs. TGF personnel thoroughly tested the TGF versions of the two scenarios prior to the DSR OT&E.

3.3 Data Collection Tools

We attempted to use the same data collection tools and techniques in the DSR Baseline as in the PVD Baseline. This included using the same questionnaires, data recording equipment, and analysis techniques. In some cases, using the identical technique was not possible, and we developed an equivalent technique. In the following subsections, we list all sources of objective and subjective data for the DSR Baseline and describe any differences between the DSR and PVD Baselines.

3.3.1 System Analysis Recording

We recorded HCS SAR tapes during each simulation run. These tapes provided data for the following metrics: operational errors, conflict alerts, halo initiations, data block positioning, aircraft under control, data entries, and data-entry errors. We also recorded DSR SAR tapes during each simulation run for use as a backup.

3.3.2 Aircraft Management Program

We recorded HCS Aircraft Management Program (AMP) tapes during each simulation run. These tapes provided data for the following metrics: average time in sector, number of arrivals, number of departures, number of overflights, number of jet aircraft, and number of propeller aircraft.

3.3.3 Target Generation Facility Recording

The TGF system automatically recorded pseudopilot actions during each simulation run onto 8-mm data tape. These recordings provided data for the number of altitude, speed, and heading changes.

3.3.4 Voice Switching and Control System

The VSCS recorded a log of the air-ground PTTs and ground-ground PTTs. These recordings provided data for the communication taskload and coordination taskload metrics.

3.3.5 Video and Audiotapes

Three low-light video cameras recorded controller activities onto Super-VHS tape. The cameras were positioned above and behind the DSR consoles so that we could see both members of the controller team. The cameras received audio input from wireless microphones worn by the controllers and from the VSCS. This provided audio recordings of air-ground, ground-ground, and non-radio communications (e.g., when a controller spoke to the other member of the controller team).

We also recorded audiotapes using the Legal Recorder system of the VSCS. These tapes recorded only air-ground and ground-ground communications. The VSCS recordings served as an audio feed for the videotapes and as a backup.

3.3.6 Questionnaires

We administered the following questionnaires during the DSR Baseline (Appendix A). When possible, these questionnaires were identical to those used in the PVD Baseline. When a questionnaire item no longer applied to the DSR, we revised or omitted the item.

a. The Background Questionnaire was completed by all controllers before the first simulation run. It collected demographic information about the controllers such as their age and experience.

b. The Post-Scenario Questionnaire was completed after each run by the controllers who worked traffic during that run. It collected controller ratings about the run such as their workload and performance. Please note that the 8-point scale used on this version of the questionnaire differs from the 7-point scale presented in the Methodology Guide (Allendoerfer & Galushka, 1999). We used the 8-point scale to be consistent with the PVD Baseline. However, we recommend that future baselines use a 7-point scale to provide consistency with the ATWIT workload ratings.

c. The Observer Evaluation Form was completed after each run by the SME observing that run. It collected SME ratings and comments about controller performance. This questionnaire has been used extensively at the Technical Center and experimentally validated (Sollenberger, Stein, & Gromelski, 1997).

d. The Observer Log was completed by an SME Observer when an unusual occurrence such as an operational error occurred during a run. SME Observers noted the time and relevant facts about the occurrence so it could be reviewed later.

e. The Final Questionnaire was completed by all controllers after the final simulation run. It collected controller ratings and comments about usability and user satisfaction with the DSR. Where necessary, we changed the wording of items to reflect differences in systems. For example, we replaced "switches" for the PVD Baseline with "on-screen controls" for the DSR Baseline.

f. The ATWIT Questionnaire was completed by all controllers after the final run. It collected validation information about the ATWIT and ensured that controllers had made their ATWIT ratings properly.

3.3.7 Workload Assessment Keypads

In the PVD Baseline, controllers made ATWIT ratings by typing a special HCS entry when a tone sounded in the control room. In the DSR Baseline, however, we administered the ATWIT using four Workload Assessment Keypads (WAKs) positioned on the DSR consoles. The WAKs provided an efficient and accurate way to administer the ATWIT and did not require hardware or software changes to the DSR.

The WAKs consisted of several numbered and lighted keys and a tone generator. The WAKs were connected to a laptop computer that controlled the timing of prompts and recorded responses. Every 4 min during each run, the WAKs emitted beeps and illuminated their lights. This prompted each participant to make a subjective workload rating from 1 (low) to 7 (high) by pressing the appropriate key. The R and D controllers made separate workload ratings. Occasionally, the SME Observers or other participants needed to remind the participants to respond. When the rating had been successfully made (or 20 sec passed), the lights extinguished. The ratings were recorded on the laptop hard disk. The WAKs provided data for the ATWIT Workload metric. We used the ATWIT Questionnaire to ensure that controllers understood the ratings they were making and the anchors of the rating scale.

Though we administered the ATWIT differently in the baseline studies, the rating scales and timing of prompts were identical. We believe that the WAKs provided a far more efficient way to collect and analyze workload ratings than the original method, and our participants found the WAKs easy to understand and use. We believe that this difference in data collection technique had no impact on the actual ratings given by our participants.

3.3.8 Pilot Test Instruments

Because controllers experienced with the DSR were available during the DSR Baseline, we used the opportunity to pilot test two data collection instruments. Neither of these instruments provided formal baseline data but may be used in future studies. The first instrument, the Keyboard Data Recorder (KDR), recorded a keystroke-by-keystroke log of each controller's data entries. The participants completed the second instrument, the DSR Keyboard Questionnaire (Appendix A), after the final run and during subsequent DSR OT&E weeks. This questionnaire collected information about areas of concern with the DSR keyboard.

3.4 Simulation Schedule and Procedure

On the Friday preceding the first baseline run, we conducted an opening briefing and informed the controllers and observers of their responsibilities during the baseline. The group discussed confidentiality and informed consent, the airspace and the traffic scenarios, operation of the WAKs, and the simulation schedule. Participants were assigned to two-person teams and assigned a participant number. The participants also completed the Background Questionnaire at this briefing.

Starting the following Monday, we conducted four simulation runs each day from 1600 hrs until 0000 hrs. We alternated between the adjacent (sectors 26 and 38) and non-adjacent (sectors 27 and 35) scenarios each run. Four participants worked traffic during each run, two serving as R controllers and two serving as D controllers. Within each team, the participants alternated between the R and D positions. We designed the simulation schedule so that no controller staffed the same position in the same sector more than once. However, an automobile accident involving several of the participants forced us to revise the schedule somewhat. Ultimately, every participant worked at least five runs with most participants working seven.

During each simulation run, the participants controlled traffic as they would at ZDC. The R controllers communicated with aircraft, issued clearances, and provided separation. The D controllers marked strips, coordinated, and assisted the R controllers as needed. The SME Observers sat behind each sector, observed controller actions, and recorded any unusual occurrences in the Observer Log.

At 4-min intervals during each run, the WAKs prompted for ATWIT workload ratings. The participants made ratings by pressing the appropriate key. After each run, the participants completed the Post-Scenario Questionnaire, and the SME Observers completed the Observer Evaluation Form. All other data sources were recorded automatically and required no action from the participants or the SME Observers.

After all runs were complete, we conducted a post-simulation briefing. At this briefing, the participants completed the Final Questionnaire and the DSR Keyboard Questionnaire. We encouraged the participants to discuss their experiences in the simulation and with the DSR. We incorporated many of their comments about improving the baseline process into the Methodology Guide (Allendoerfer & Galushka, 1999).

4. Results

Whenever possible, we used the identical data reduction and analysis (DR&A) tools and procedures as the PVD Baseline. However, because the DSR Baseline used a different simulation platform, a different communications platform, and some different data collection tools, some metrics required the development of new Data Reduction & Analysis (DR&A) procedures. In these cases, we developed the new procedures so that they followed the originals.

We reduced the HCS SAR tapes using the Data Analysis and Reduction Tool (DART) and the AMP tapes using the Offline Aircraft Management Program. We further processed the output of these tools to organize the data and make it easier to interpret. These tapes provided data for most of the objective metrics. We reduced the questionnaire data by manually entering the responses into a spreadsheet. After data entry was complete, we thoroughly reviewed the data. We entered handwritten comments into a word processor and edited for spelling and grammar. We reduced data for the number of altitude, speed, and heading changes from 8 mm tape using DR&A routines developed by the TGF. We recorded ATWIT workload ratings directly into a database file. The VSCS software counted the number of air-ground and ground-ground communications electronically .

Appendix B provides the complete DSR Baseline data. The format of this appendix closely follows the format used in the PVD Baseline. Data are reported at one or more levels of detail:

a. Overall Level: This level provides data aggregated across all intervals, sectors, and runs. We report data from the Final Questionnaire only at this level because this questionnaire was administered only once after all simulation runs were complete. Data for the traffic scenario characteristics are not reported at this level because it is not meaningful to average these data across sectors.

b. Sector Level: This level provides information about individual sectors aggregated across intervals and runs. We provide the means and standard deviations for each sector. Note that sectors 26 and 38 used 60-min scenarios, whereas sectors 27 and 35 used 90 min scenarios. Because of this, metrics based on totals such as the number of data entries will usually be higher in sectors 27 and 35.

c. Interval Level: This level provides information about individual 12-min intervals aggregated across runs. This level best demonstrates changes resulting from changes in the traffic volume and complexity. The means and standard deviations for each sector and each interval are provided. Note that sectors 26 and 38 have 5 intervals, whereas sectors 27 and 35 have 7.

5. PVD and DSR Baseline Comparison

The main purpose for conducting the PVD and DSR Baselines was to directly compare the systems. In particular, we wished to assess the effects of the DSR on safety, capacity, performance, workload, and usability. The following sections compare the data from the baselines and discuss the implications for the DSR.

Operational input is crucial to understanding the causes and implications of any differences between systems. To provide this input, we assembled an Operational Review Team consisting of

a. engineering research psychologists who were involved in the data collection and analysis;

b. the National Air Traffic Controllers Association (NATCA) representative to the DSR OT&E and Baseline;

c. the FAA Air Traffic Supervisors Committee (SUPCOM) representative to the DSR OT&E and Baseline;

d. two FPL controllers from ZDC who had served as participants in both baselines; and

e. technical personnel from the TGF, HCS, PVD, and DSR facilities at the Technical Center, as needed.

The goals of the team were

a. to compare the two systems along the five operational constructs and identify differences,

b. to identify potential causes for the differences,

c. to assess the implications of the differences,

d. to identify aspects of the DSR that merit further study and improvement, and

e. to recommend ways in which the baseline process could be improved.

We led the team through a briefing showing graphs comparing the PVD and DSR Baseline data. We encouraged the team members to ask questions, discuss results, and request additional data analyses. This was an iterative process that took nearly 2 weeks to complete. The results of the review are presented in the following sections.

The graphs usually compare the systems at the sector level but, when appropriate and informative, we provide graphs showing the 12-minute interval level. The team concluded that methodological differences between the baselines had invalidated the comparison for some metrics. In these cases, the team agreed to exclude the metric from the comparison, and we discuss the exclusion rationale in the following subsections.

5.1 Safety

5.1.1 Operational Errors

One operational error was initially identified for the DSR Baseline using automated DR&A tools. However, because no errors had been recorded by the SME Observers, the team reviewed video and audiotapes of the error to determine whether it had resulted from a genuine controller mistake or was an artifact of the simulation. The team concluded that the error resulted from a pseudopilot mistake and agreed that it was not a genuine operational error. As a result, the team concluded that no operational errors occurred in either baseline.

However, the controllers on the team believed that the traffic scenarios used in the baselines were not complex enough to show differences in the number of operational errors. They based this conclusion on their observations that genuine operational errors occurred during other OT&E activities where a higher level of traffic volume and complexity was used. The team recommended increasing the traffic volume in future baselines to study operational errors more closely.

5.1.2 Conflict Alerts

The review team raised two concerns about this metric and agreed to exclude it from the comparison. First, despite the high number of alerts (more than one per run), the team members did not remember this many alerts occurring. They suspected that most alerts resulted from the techniques used by the simulation platforms to initiate simulated aircraft. For example, two aircraft might be created already in an alert or near-alert situation. Though the scenarios were designed so that all aircraft were separated by the time they reached the operational sectors, it is possible that conflict alerts persisted for several sweeps. In this case, the conflict alert was not caused by any action or inaction by a participant and should not be counted as a genuine alert. Unfortunately, an effort to review each conflict alert from the videotapes and separate genuine alerts from spurious ones was not feasible during the review.

Second, the controllers on the team questioned whether even genuine conflict alerts would provide information about safety. They explained that some controllers "control by conflict alert" whereby they allow aircraft to fly at separations close enough to activate the conflict alert but not close enough to cause an operational error. These controllers, they said, use conflict alert as a separation tool rather than as a warning. For these reasons, the team agreed to exclude this metric from the comparison.

5.1.3 Halo Initiations

As shown in Figure 1, in sectors 26, 27, and 35, the participants initiated the halo more frequently in the DSR Baseline. In sector 38, the participants initiated the halo slightly more in the PVD Baseline. The team concluded that the differences shown here resulted from two factors. First, the controllers on the review team explained that using the halo requires only a single entry in the DSR, whereas two are required in the PVD. This made the halo quicker, easier, and more desirable to use. Second, the controllers explained that the vector lines were more difficult to use in the DSR, causing participants to reduce their use of the vector lines in favor of the halo. Unfortunately, no data about vector-line were recorded during the PVD Baseline, so no analysis of this insight could be performed. The team concluded that this difference in halo usage did not result from a difference in ability to separate aircraft but rather on a difference in the computer-human interfaces (CHIs) of the systems.

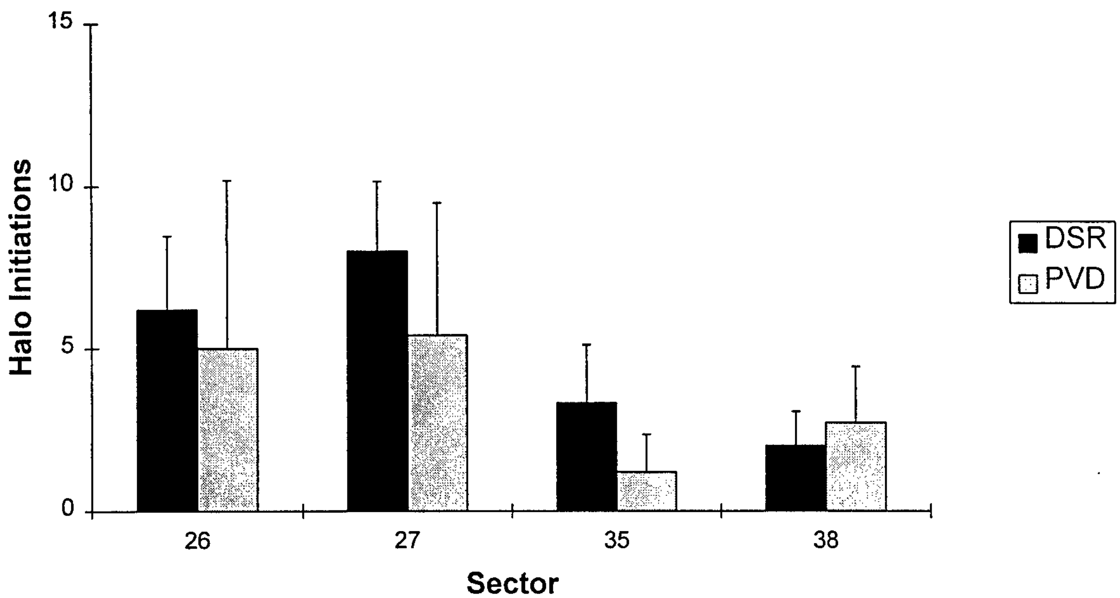

Figure 1. Number of halo initiations for each sector, averaged across runs.

5.1.4 Data Block Positioning

As shown in Figure 2, participants in every sector repositioned the data blocks more frequently in the DSR Baseline than the PVD Baseline. The team concluded that this difference resulted from a difference in readability when data blocks overlap. The controllers on the review team explained that when two data blocks overlap on the PVD, the overlapping characters are still somewhat readable unless the characters are almost entirely overlapped. However, when two data blocks overlap on the DSR, a much smaller amount of overlap is necessary to render the characters unreadable. The participants referred to this effect as "the green blob." It requires controllers to be extra vigilant in their data block positioning to maintain readability. The controllers on the team believed it increased their workload but did not reduce safety because of the low traffic complexity of the scenarios. The team agreed that the increase in data block positioning did not result from closer aircraft proximity but rather from a problem with the DSR CHI. The team agreed that this problem is serious enough to warrant further study and possible improvement via the Pre-planned Product Improvement (P^3I) process.

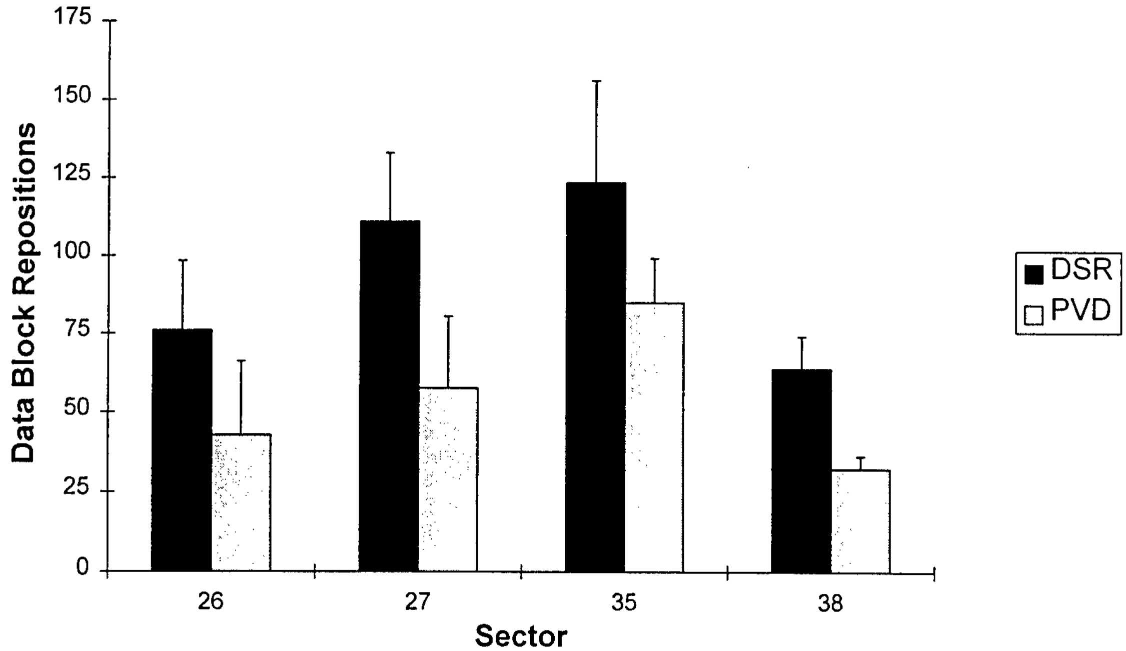

Figure 2. Number of data block positioning actions for each sector, averaged across runs.

5.1.5 Other Safety-Critical Issues

Though no safety-critical issues were reported on the Observer Logs in either baseline, the SMEs on the review team identified several aspects of the DSR that warrant further study. First, they were concerned about increased heads-down time required by the new DSR keyboards, particularly at the R position. Second, the review team members were concerned about the impact on safety of increased data entry errors, particularly at high traffic volumes. Third, they were concerned about the impact of data block overlap, particularly in high-volume sectors where aircraft are tightly packed and where being able to read altitude information is especially important. Fourth, they were concerned with the length and configuration of flight strip bays at the D position. The several short bays on the D position may require the D controller to order, organize, and purge strips more frequently than when using the two longer bays provided by the PVD console.

5.2 Capacity

5.2.1 Aircraft Under Control

As shown in Figures 3a through d, the number of aircraft under control during each 12-minute interval varied only slightly between baselines. In both baselines, the progression of the traffic scenario is reflected in the changing number of aircraft under control in each interval. Both baselines show patterns with very similar shapes, demonstrating that not only did the number of aircraft remain constant, the traffic pattern also remained constant. The team agreed that the DSR did not affect the number of aircraft that could be controlled in these scenarios.

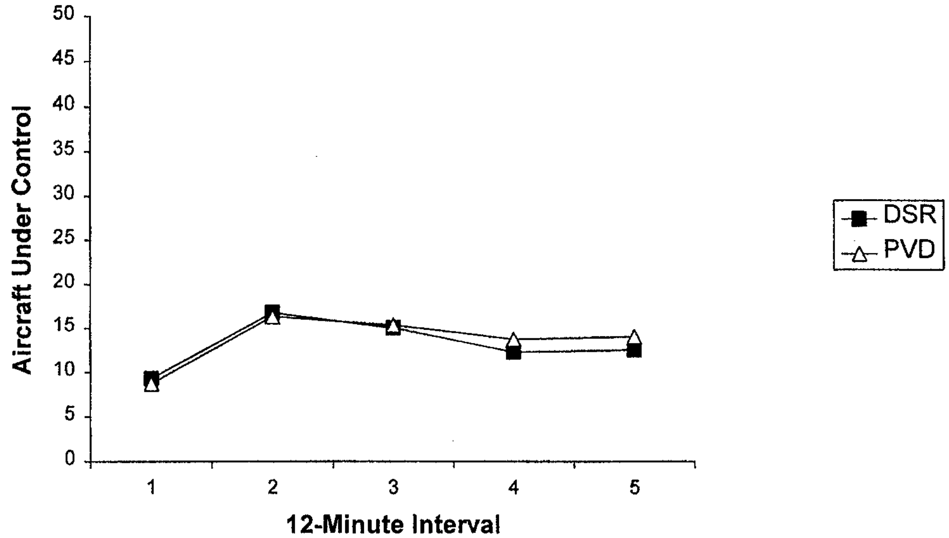

Figure 3a. Number of aircraft under control for Sector 26, averaged across runs.

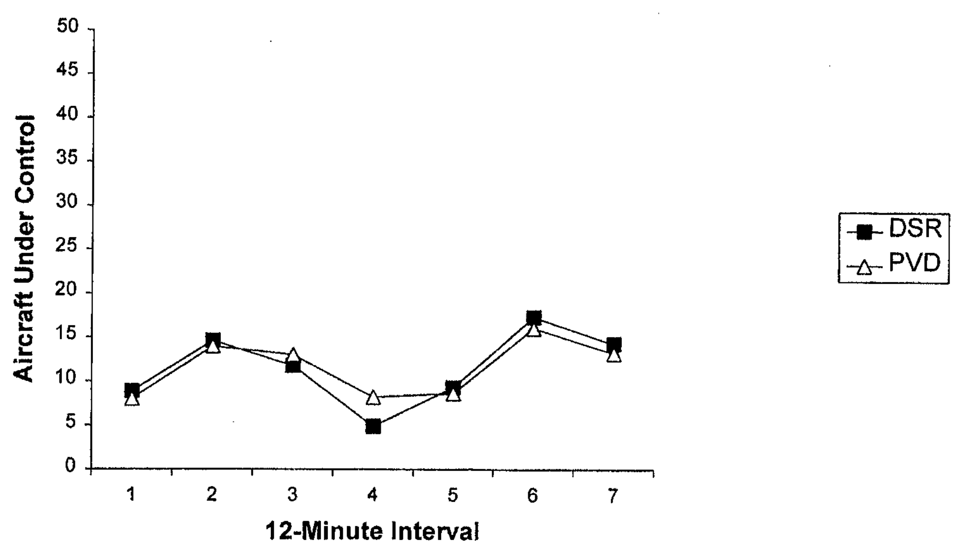

Figure 3b. Number of aircraft under control for Sector 27, averaged across runs.

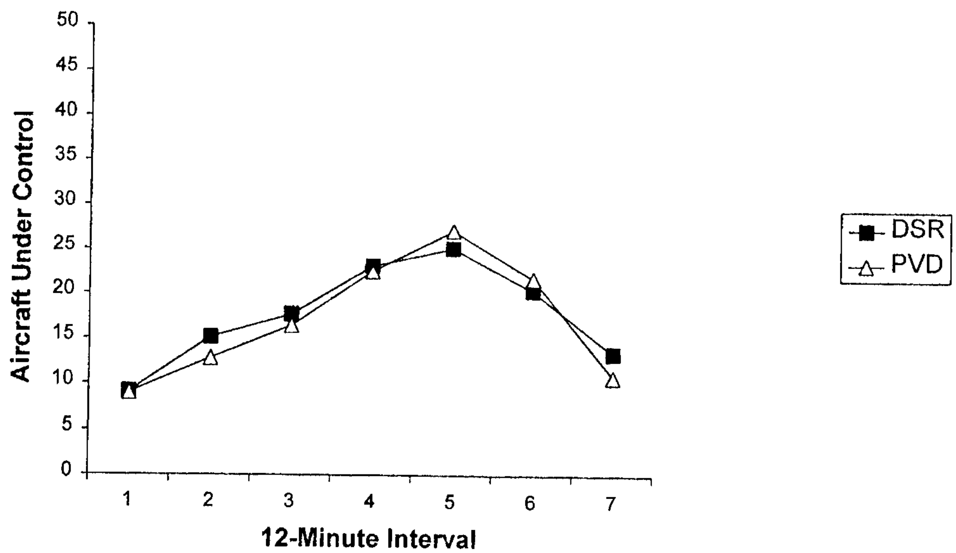

Figure 3c. Number of aircraft under control for Sector 35, averaged across runs.

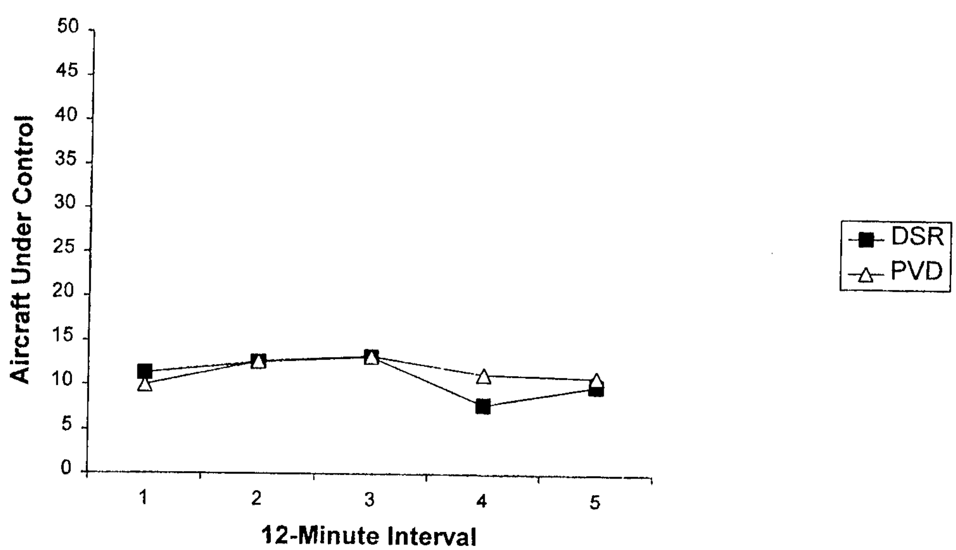

Figure 3d. Number of aircraft under control for Sector 38, averaged across runs.

5.2.2 Time in Sector

The review team uncovered two inconsistencies between the baselines that affected this metric and agreed to exclude it from the comparison. First, the DYSIM and TGF simulation platforms used different aircraft performance models, so the same aircraft may climb, descend, and turn at different rates on the two simulation platforms. If a difference did appear between the systems on this metric, it would not be possible to attribute it to the system or the simulation platform. Second, controller pseudopilots in the PVD Baseline may have adjusted the aircraft speeds according to their own knowledge of aircraft capabilities. TGF pseudopilots in the DSR Baseline did not make discretionary speed modifications. The team agreed that too many confounds existed for this metric and decided not to include it in the comparison.[1]

5.3 Performance

5.3.1 Overall Data Entries

As shown in Figure 4a, R controllers made more data entries in the DSR Baseline than in the PVD Baseline. The magnitude of this difference varied by sector. The review team attributed this difference to four factors. First, data blocks are positioned via an HCS entry. The increased need to keep data blocks separated (see Section 5.1.4) increased the overall number of data entries made in the DSR Baseline. Second, an increase in data entry errors necessarily results in more data entries because every incorrect entry requires subsequent re-entry of the original message. Because an increase in data-entry errors for the DSR was also found (see Section 5.3.2), the review team concluded that some of the increase in data entries was due to these errors. Third, D controllers appeared to be less involved during the DSR Baseline (see Section 5.4), and R controllers may have made entries normally entered by the D controllers. Fourth, because the halo is initiated via an HCS entry, if controllers shifted away from vector lines in favor of the halo, more data entries would result in the DSR.

As shown in Figure 4b, D controllers made more data entries in the PVD Baseline than the DSR Baseline in sectors 26, 27, and 35. In sector 38, D controllers made about the same number of entries in both baselines. The review team attributed this difference to reduced involvement of the D controllers in the DSR Baseline (see Section 5.4) and the increased between-sector coordination requirements in the PVD Baseline (see Section 5.4.4).

[1] This inconsistency calls into question results reported in a comparison of the PVD Baseline to the Eurocontrol ODID IV experimental ATC system (Keegan, Skiles, Krois, & Merkle, 1996; Krois & Marsden, 1997). In that study, the ODID IV allowed aircraft in sector 26 to traverse the sector in 1.4 min less time than in the PVD Baseline. We recommend re-examining their data to ensure that Eurocontrol used equivalent aircraft performance models to the DYSIM.

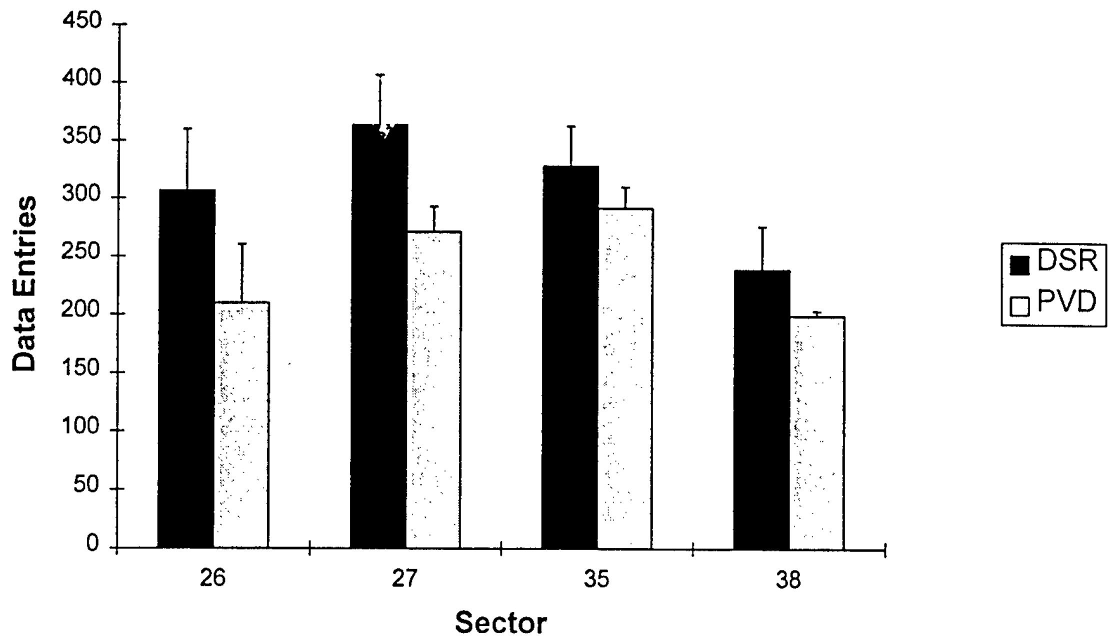

Figure 4a. Number of data entries made by the R controllers for each sector, averaged across runs.

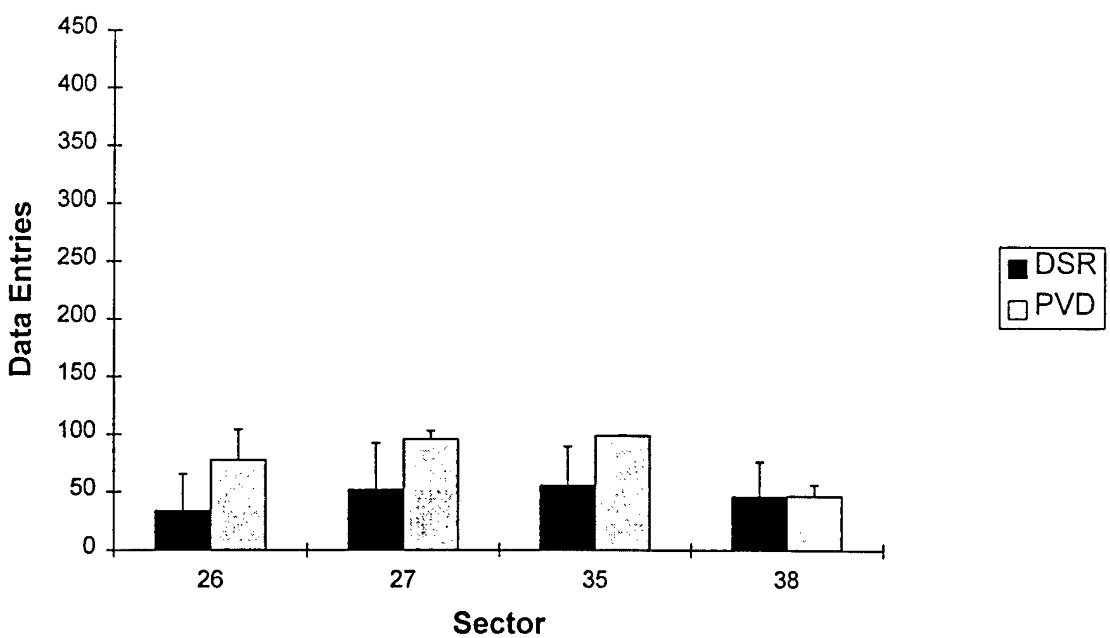

Figure 4b. Number of data entries made by the D controller for each sector, averaged across runs.

5.3.2 Specific Data Entry Types

Because there are at least 40 different HCS data entry types, we compared only the most common and important functions. We discuss the important differences here. Data entry types that are not discussed did not show consistent or meaningful differences between systems. Further comparisons can be made using the values listed in the appendixes of the PVD and DSR Baselines.

5.3.2.1 FP and SR

In the PVD Baseline, the participants made many FP (allows the entry of flight plan data) and SR (outputs a flight progress strip) entries, whereas the participants made almost none in the DSR Baseline. This difference resulted from differences in the simulation platforms. In DYSIM, flight plans and accompanying flight strips are not always generated automatically and must frequently be entered by the controller. In TGF, these actions are not necessary because the simulation platform correctly creates the flight plan, sends it to the HCS, and requests a flight progress strip. As a result, these entries were not made in the DSR Baseline.

5.3.2.2 QN and QZ

The QN and QZ data entry types are used to accept handoffs, initiate handoffs, force data blocks, and change data block positions. QZ is also used to assign altitudes. The two functions are largely redundant and interchangeable. If the controller presses the QN/QZ (None) Quick Action Key (QAK), the entry is logged as a QZ entry. If the controller does not press the QAK and instead makes an implied entry, the entry is logged as a QN entry. Participants in both baselines used the QN version of these functions much more frequently than the QZ version. However, controllers in the PVD Baseline made somewhat more QN entries in sectors 27 and many more in sectors 35 and 38. Without reviewing each QN entry individually, it is difficult to determine why this occurred. However, the team suspected that this resulted from the simulation platforms. The DYSIM required data blocks to be forced to the ghost sector when the aircraft were handed off. In sector 26, the aircraft were handed off to sector 38 and did not require a data block force. This requirement did not exist with the TGF, so these additional entries were not made.

5.3.2.3 QP

The QP data entry type is used for point outs, to request and suppress data blocks, to move lists, and to activate the halo. As discussed in Section 5.1.3, controllers in the DSR Baseline used the halo more often than in the PVD Baseline because the vector-line function was difficult to use.

5.3.3 Data Entry Errors

As shown in Figure 5a, R controllers made more data entry errors in the DSR Baseline. The controllers on the review team attributed this difference to several problems they experienced with the DSR keyboards:

a. In the PVD, the QAKs are located on the console. In the DSR, the QAKs are located on the top two rows of the keyboard. This places them in closer proximity to other keys and items placed on the work surface such as flight strip holders and documentation. As a result, in the DSR, participants often accidentally pressed QAKs, whereas they very seldom accidentally pressed QAKs in the PVD.

b. The QAKs on the DSR keyboard are organized and labeled differently than the QAKs on the PVD. As a result, participants sometimes pressed the wrong QAK or had to spend more time locating and identifying the correct QAK.

c. The Clear key is located in the upper right corner of the main keyboard group where the Backspace key is traditionally located on a standard PC keyboard. The Backspace key is located next to the Right Shift key where the question mark is traditionally located. As a result, when participants meant to press Backspace, they often cleared the entry by mistake and then unknowingly sent an incomplete entry to the HCS.

d. The numeric keypad includes a Space key in a location that is dissimilar to both a standard computer keyboard and the PVD keyboard. The participants reported that they often accidentally pressed this key while making numeric entries. This resulted in incorrect entry syntax.

e. The zero key on the DSR numeric keypad is located between two other keys. This is dissimilar from the PVD keyboard where the zero does not have keys on either side. The participants reported that they accidentally pressed these keys when trying to press the zero key. This resulted in incorrect entry syntax.

f. The DSR keys require less force to press than the PVD keys. The participants reported that they frequently made errors when they used the amount of force to which they were accustomed and inadvertently pressed multiple keys. This resulted in incorrect entry syntax.

As shown in Figure 5b, D controllers in Sectors 26 and 27 made more data entry errors in the PVD Baseline. In Sector 35, they made about the same number of errors in both baselines. In Sector 38, they made more errors in the DSR Baseline. This inconsistent pattern should be compared to the consistent pattern found among the R controllers. The review team attributed this variable pattern to the general lack of involvement of the D controllers in the DSR Baseline and to the relatively small number of data entry errors made (around 15 per run).

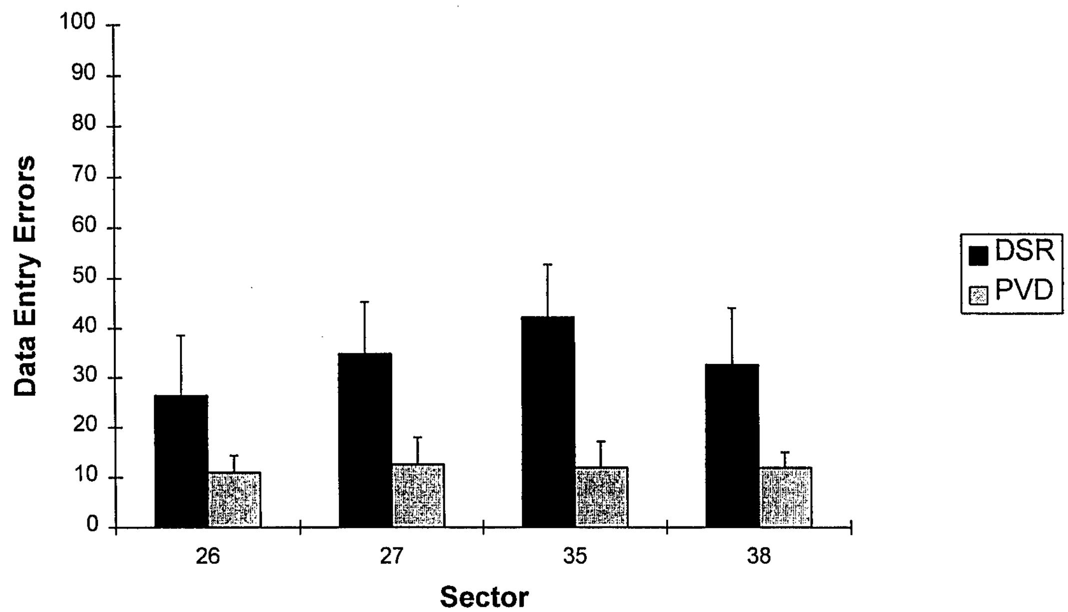

Figure 5a. Number of data entry errors made by the R controllers for each sector, averaged across runs.

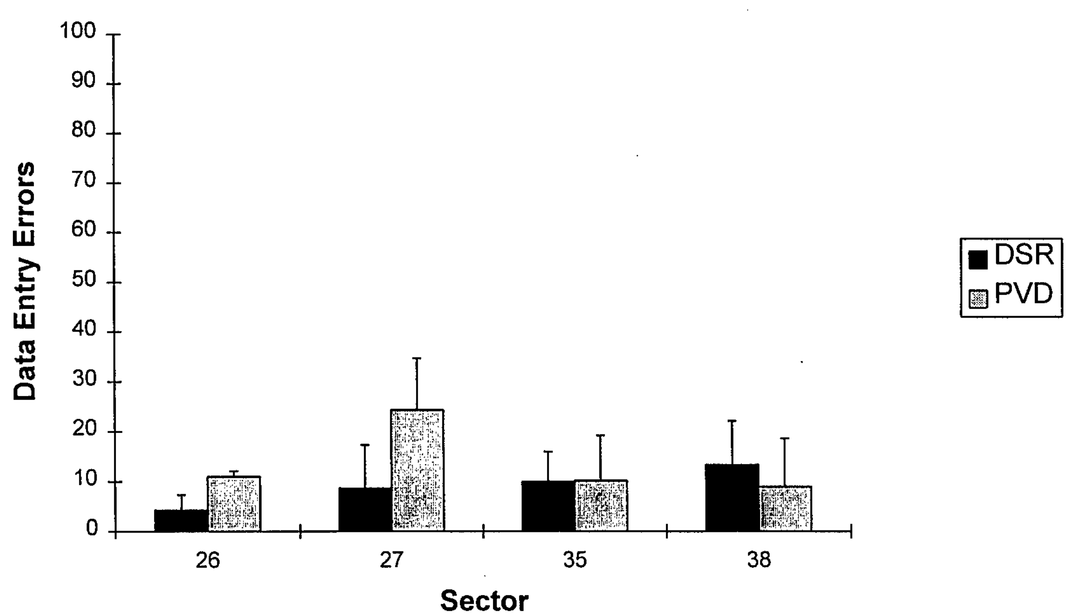

Figure 5b. Number of data entry errors made by the D controllers for each sector, averaged across runs.

5.3.4 Number of Altitude, Speed, and Heading Changes

The review team uncovered several inconsistencies between the baselines that affected this metric. The team agreed that these inconsistencies were serious enough to warrant exclusion of this metric from this comparison. We do believe that the DSR Baseline values for this metric are valid. If the TGF and the same analysis techniques are used in a future baseline, the data reported here can be validly compared.

First, at the time of the PVD Baseline, Letters of Agreement (LOAs) in place at ZDC required that Raleigh-Durham International Airport (RDU) departures be given an interim altitude of 10,000 ft before being cleared to a higher altitude. At the time of the DSR Baseline, however, this restriction had been lifted at ZDC. This LOA change was unknown to us prior to the DSR Baseline, and we made no effort to ensure the participants operated under the original LOAs. As a result, participants in the DSR Baseline controlled without the altitude restriction. This resulted in fewer altitude changes for RDU departures.

Second, LOAs did not allow aircraft at ZDC to exceed 250 knots below 10,000 ft. Because the DYSIM requires that simulated aircraft begin at full speed, the controller-pseudopilots had to slow some aircraft to 250 knots and then change the speed to full speed when the aircraft reached 10,000 ft. The TGF, however, can model more complex aircraft behavior, and the simulator could follow this restriction automatically. As a result, the pseudopilots did not make speed changes to slow down then speed up aircraft. This resulted in two additional speed changes in the PVD Baseline for every departure aircraft.

Third, in the PVD Baseline, controllers staffed the ghost sector. As a result, participants used standard operating procedures to handoff climbing aircraft, particularly RDU departures. In the DSR Baseline, however, the ghost sector was staffed by simulation support staff who were unfamiliar with ZDC LOAs and procedures. Participants could (and did, according to the controllers on the review team) climb aircraft out of their sector at a rate that would violate ZDC LOAs because the ghost sector was not knowledgeable enough to refuse the handoff. This resulted in fewer altitude changes issued to these aircraft.

We believe that the DSR Baseline values are valid in and of themselves. If the TGF, the same LOAs, and the same analysis techniques are used in a future baseline, the DSR data are suitable for other comparison.

These and other inconsistencies may also affect the conclusions that can be drawn from the comparison of the PVD and ODID baselines (Krois & Marsden, 1997; Skiles, Graham, Marsden, & Krois, 1997). The three baselines used different simulation platforms, and these platforms may differ in the amount of entries required from the pseudopilots to create realistic aircraft behavior. For example, a simple simulator might require several separate heading commands to create a realistic holding pattern. More advanced simulators are able to create a holding pattern with a single entry. Because the number of pseudopilot entries is the basis for the number of altitude, speed, and heading changes metric, the cause of a difference for this metric is unclear when compared across different simulation platforms. In future studies where this metric must be compared across simulation platforms, researchers should either use another method to calculate the metric (e.g., counting based on audio recordings) or develop a method for compensating for the differences between platforms.

5.3.5 Self-Assessments of Performance

We collected this metric from two items on the Post-Scenario Questionnaire that asked participants to rate the quality of their ATC services during the simulation. As shown in Figures 6a and b, participants rated themselves similarly in both baselines on the "How well did you control traffic during this problem?" item. In one case, sector 26 for the D controllers, the participants rated themselves as controlling traffic better in the DSR Baseline than in the PVD Baseline by more than a full rating point. Due to the small sample size in the PVD Baseline, one unusually low rating on one run was able to lower the group average enough to show this difference. The review team agreed that the data for this metric show no operationally meaningful difference between the systems.

As shown in Figures 7a and b, participants rated themselves fairly similarly in both baselines on the "How good do you think your air traffic control services were from a pilot's point of view?" item. This item, however, particularly for the D controllers, showed more variability than the previous item. In sector 26, D controllers gave themselves higher ratings in the DSR, whereas, in sector 27, they gave themselves higher ratings in the PVD. Some members of the review team questioned the utility and validity of this metric because it asked participants to rate their performance from someone else's point of view, a judgment that controllers are not accustomed to making. If participants were uncomfortable or confused by the questionnaire item, this may help explain the inconsistent ratings.

5.3.6 Observer Assessments of Performance

Due to changes in the Observer Evaluation Form between the PVD Baseline and the DSR Baseline, the Providing Control Information and Communicating items on the form could not be compared between baselines. These changes resulted from many validation activities conducted by Sollenberger, Stein, and Gromelski (1997) that improved the questionnaire in the interim between the baselines. Comparisons between items that remained fairly similar between baselines are shown in Figures 8a and b. In general, SME Observers rated participants higher on the PVD Baseline than the DSR Baseline. In every case, this difference was one rating point or less. Because the differences were small and a single obvious cause was not identified, the team developed several rationales for these differences. First, even though the participants were relatively experienced users of the DSR, they still had substantially less experience with it than the PVD. This lack of experience may have reduced their performance somewhat in the judgment of the SME Observers. Second, one of the SME Observers changed between the baselines. It is possible that the new observer had slightly higher criteria for his ratings and tended to give lower ratings.

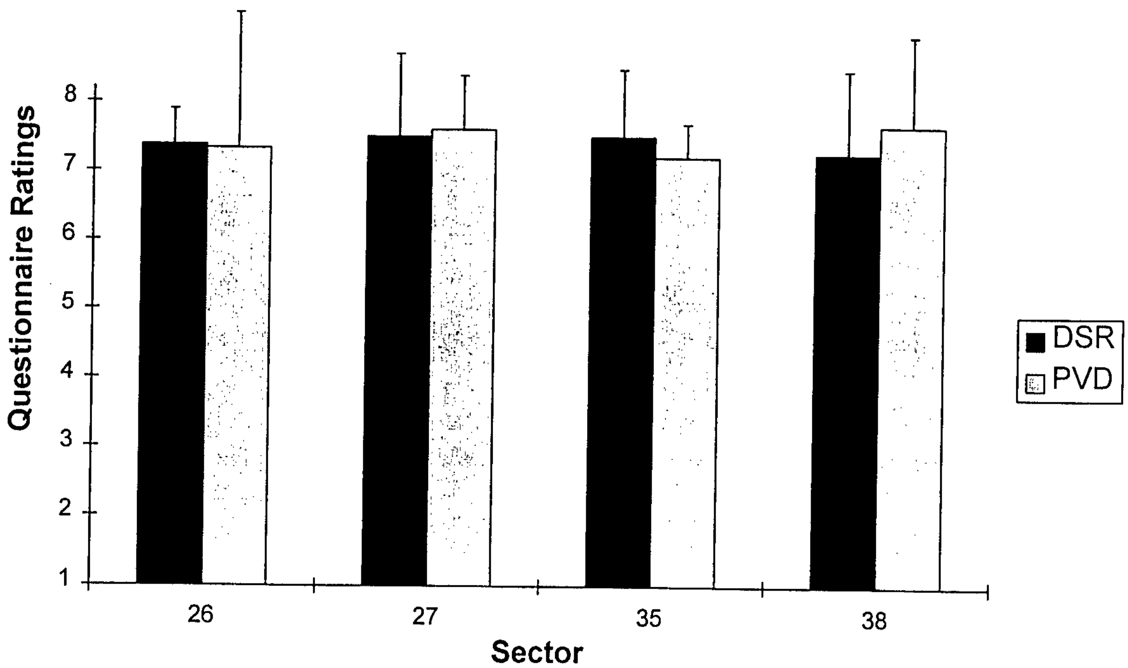

Figure 6a. Ratings given by R controllers on the "How well did you control traffic during this problem?" questionnaire item for each sector, averaged across runs.

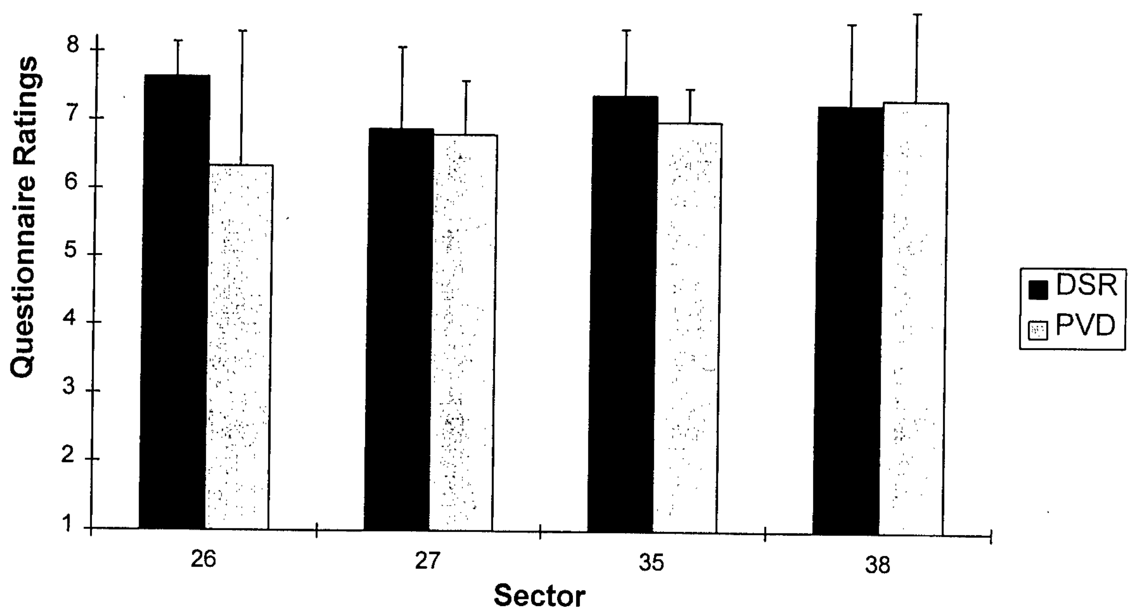

Figure 6b. Ratings given by D controllers on the "How well did you control traffic during this problem?" questionnaire item for each sector, averaged across runs.

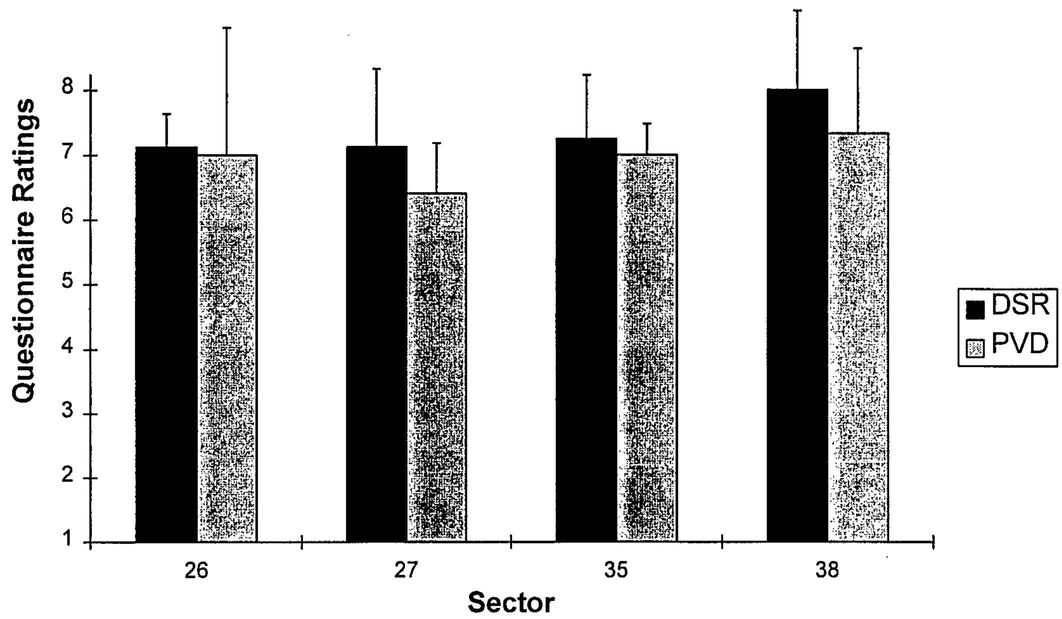

Figure 7a. Ratings given by R controllers on the "How good do you think your air traffic control services were from a pilot's point of view?" questionnaire item for each sector, averaged across runs.

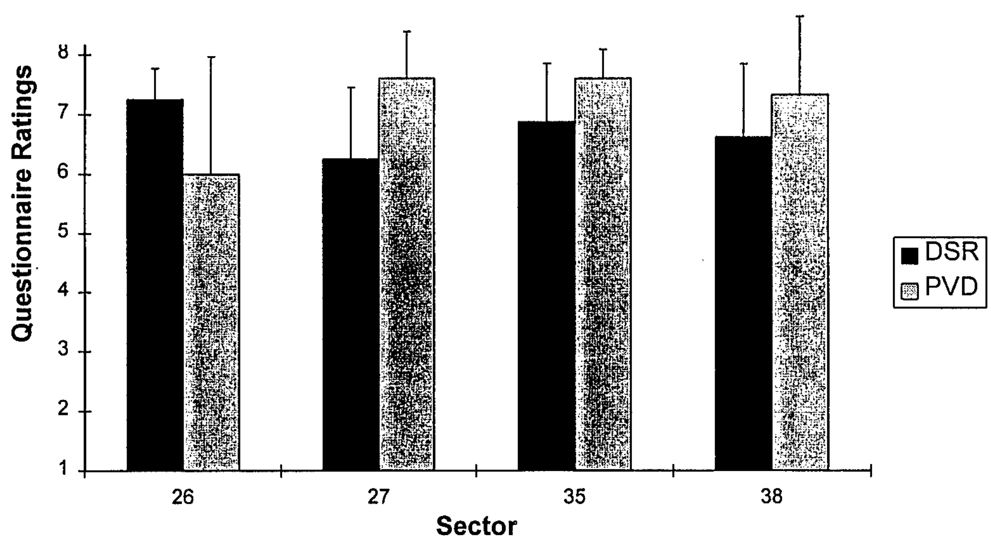

Figure 7b. Ratings given by D controllers on the "How good do you think your air traffic control services were from a pilot's point of view?" questionnaire item for each sector, averaged across runs.

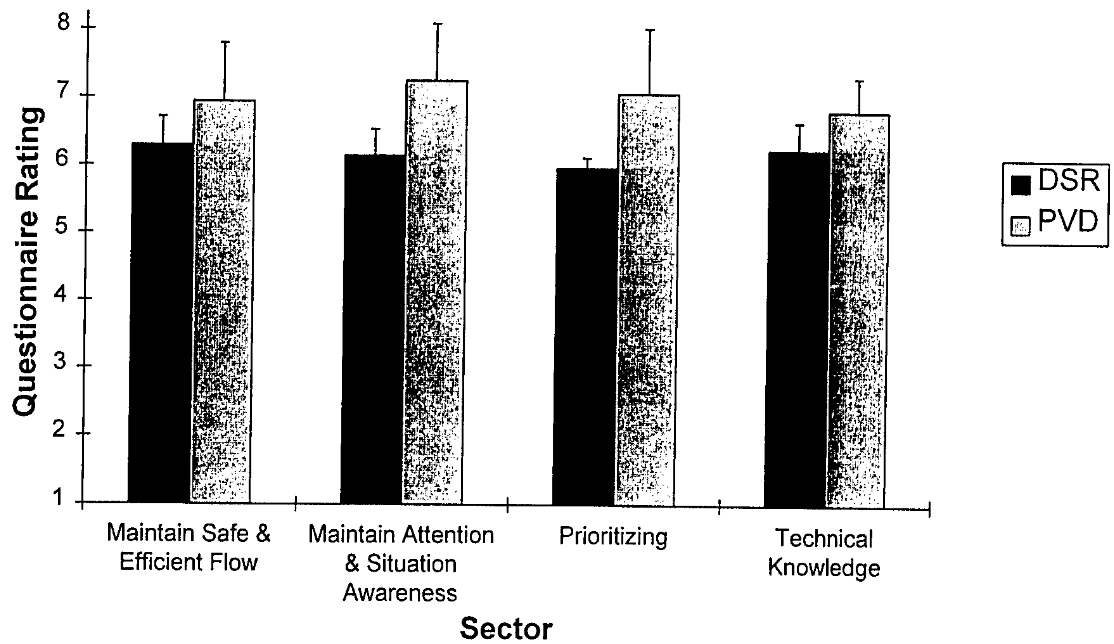

Figure 8a. Ratings given by SME Observers on the Observer Rating Form for controllers working the R position, averaged across runs.

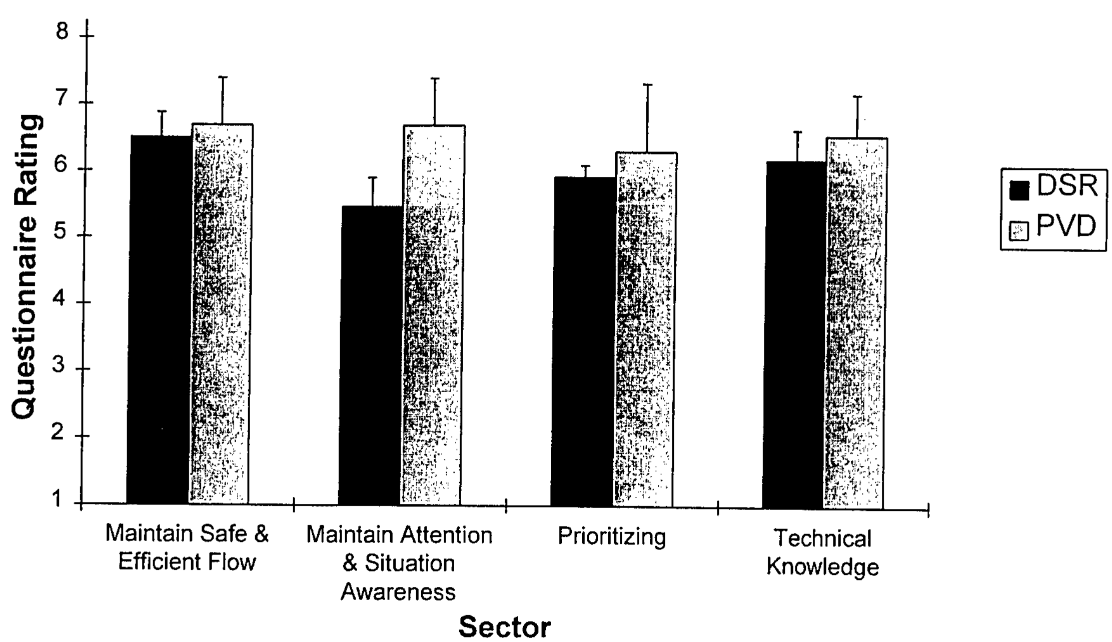

Figure 8b. Ratings given by SME Observers on the Observer Rating Form for controllers working the D position, averaged across runs.

5.4 Workload

In general, the participants rated their workload as low to moderate on the ATWIT and Post-Scenario metrics. This overall level was lower than we intended and lower than would be expected on a 90th percentile day. Some possible explanations for this are included in the sections below.

5.4.1 ATWIT Workload

As shown in Figures 9a and b, participants in the DSR Baseline rated their workload as equal to or somewhat lower than the PVD Baseline. The largest difference, 1 rating point, occurred for the R controllers in Sector 26. The team identified several rationales for these small differences in workload.

a) Participants served as their own pseudopilots in the PVD Baseline and may have taken more discretionary actions (e.g., speed changes) during the course of the simulation. These discretionary actions may have increased controller workload and changed the nature of the traffic scenario such that the scenario seemed more difficult.

b) Due to different flight strip print parameters, more strips were printed earlier in the PVD Baseline simulation runs, increasing the perception of urgency and pending traffic. Though the actual number of planes in the two studies were identical, participants in the PVD Baseline may have believed that more planes were on their way and adjusted their workload ratings as a result.

c) As discussed in Section 5.3.4, some LOAs had changed since the PVD baseline, and the participants handled RDU departures differently. This change in procedures may have reduced the number of actions required to control traffic.

d) The controller staffing the ghost sector was not a ZDC controller and did not know when it was appropriate to approve or reject a point out or handoff. As a result, some participants chose to stop making verbal coordination with the ghost sector thereby eliminating tasks associated with contacting the ghost sector through the VSCS.

Data from the ATWIT Questionnaire revealed that all participants understood the ATWIT scale and were making ratings appropriately. All participants reported that they correctly made low ratings when their workload was low and high ratings when their workload was high.

5.4.2 Post-Run Workload

As shown in Figures 10a and b, the participants rated their workload on the Post-Scenario Questionnaire in a similar pattern to the ATWIT Workload ratings. In general, participants rated their workload lower in the DSR Baseline than the PVD Baseline. The review team agreed that that this difference was due to the factors identified for the ATWIT Workload Ratings.

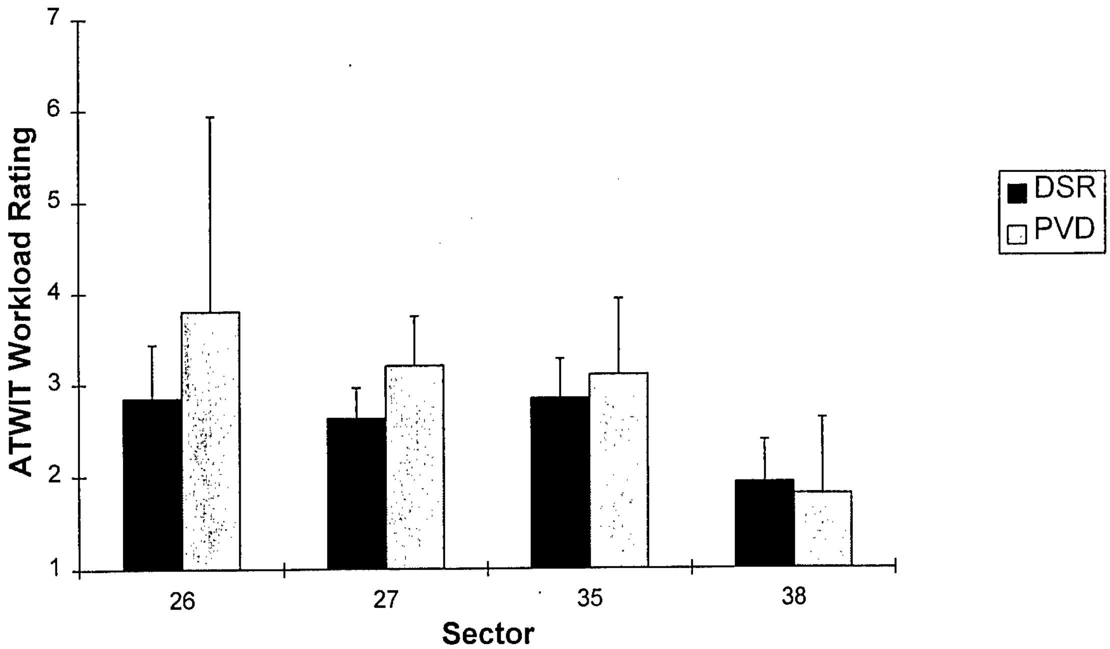

Figure 9a. ATWIT workload ratings given by R controllers for each sector, averaged across runs.

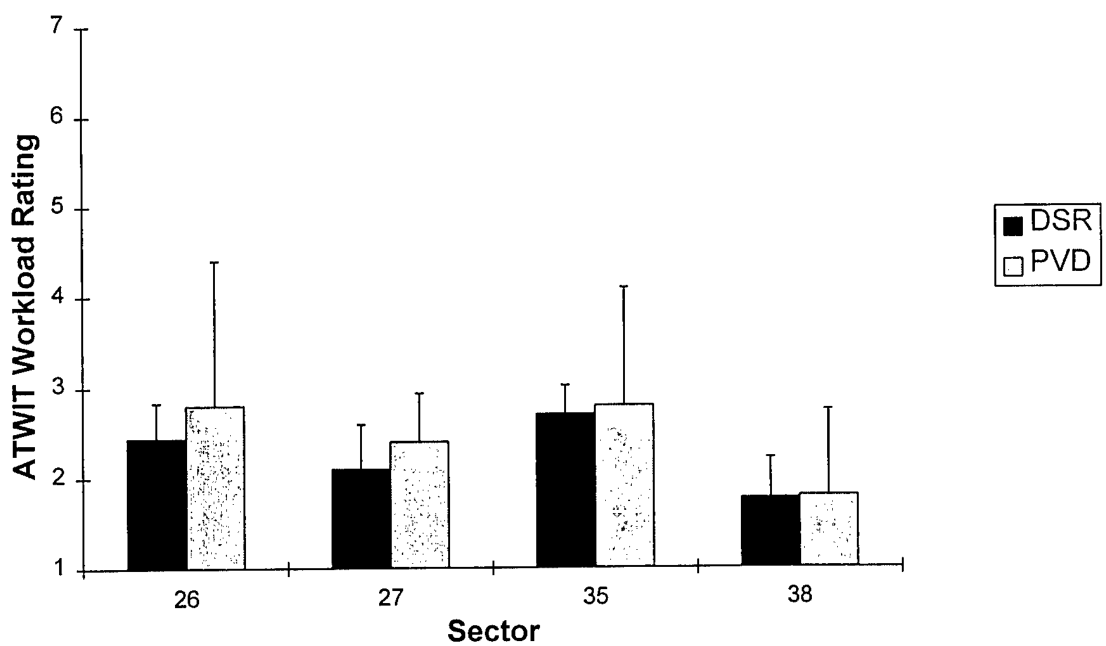

Figure 9b. ATWIT workload ratings given by D controllers for each sector, averaged across runs.

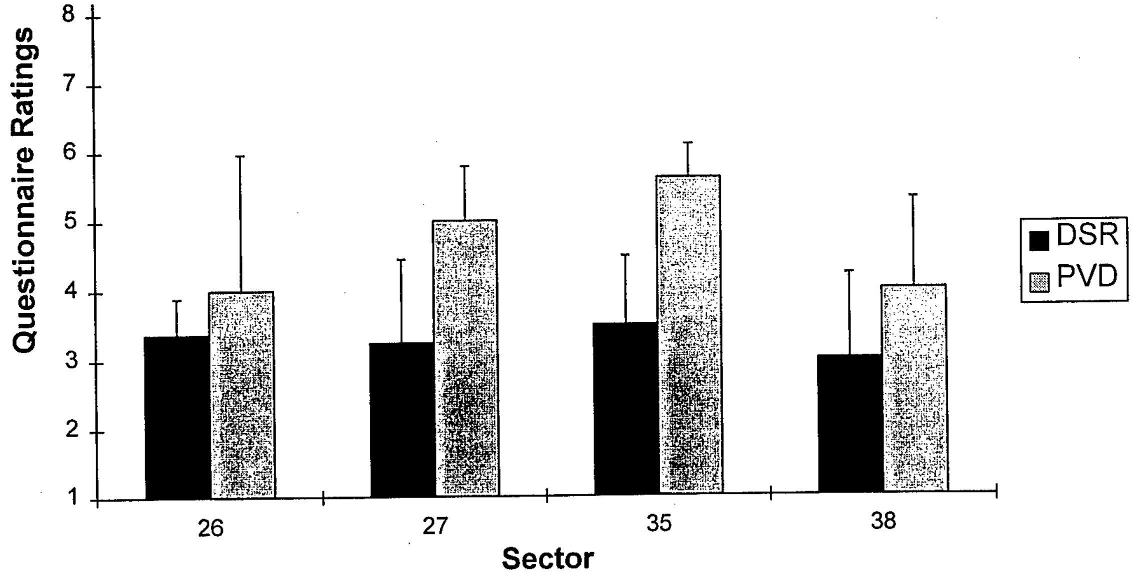

Figure 10a. Post-Scenario Workload ratings given by R controllers for each sector, averaged across runs.

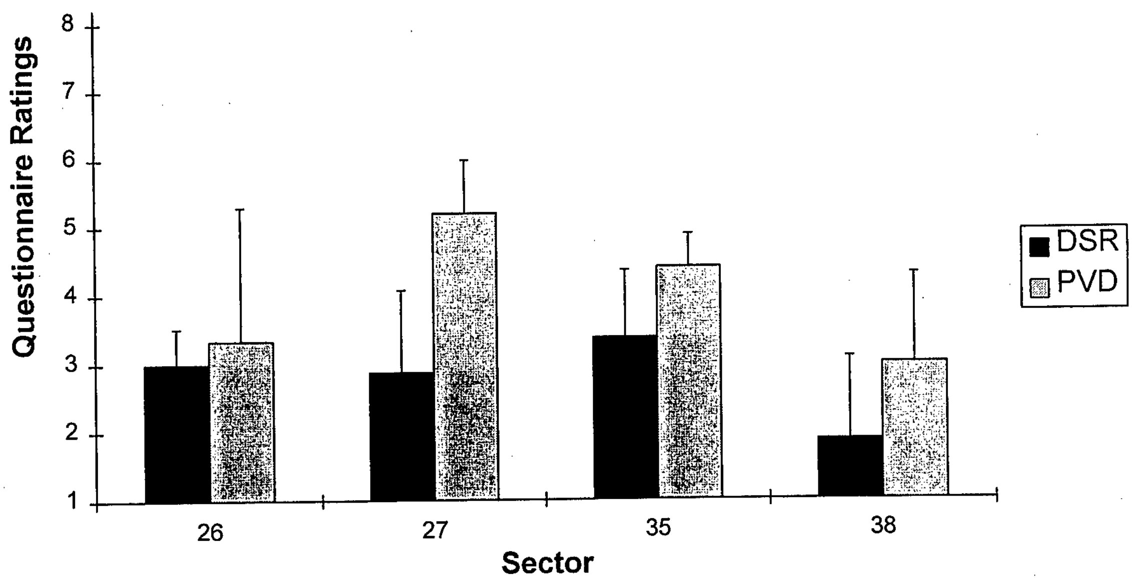

Figure 10b. Post-Scenario Workload ratings given by D controllers for each sector, averaged across runs.

5.4.3 Communication Taskload

The review team uncovered an inconsistency between the baselines that warranted exclusion of this metric from the comparison. In the PVD Baseline, manually reviewing audiotapes and videotapes and counting PTTs by hand calculated the number of PTTs, whereas in the DSR Baseline, the VSCS counted air-ground PTTs automatically. The VSCS counted a PTT each time the controllers activated their microphones. However, in the PVD Baseline, if the controllers activated their microphones but did not speak, nothing would have been recorded on the audiotape and no PTT would have been counted. Because it was not feasible to review all the communications in the DSR Baseline during the review period, this metric was excluded from this comparison. We believe that the DSR Baseline values are valid and, if the VSCS and the same analysis techniques are used in future baselines, the data are suitable for other comparisons.

5.4.4 Coordination Taskload

The review team uncovered two inconsistencies between the baselines that warranted exclusion of this metric from the comparison. First, in the PVD Baseline, a ZDC controller staffed the ghost sector. As a result, handoffs and point outs were approved only if they would have been approved in the field. However, in the DSR Baseline, a simulation support specialist unfamiliar with ZDC LOAs and procedures staffed the ghost sector. As a result, all handoffs and point outs were approved, regardless of whether they would have been in the field. As the simulation continued, the participants stopped coordinating with the ghost sector because they felt it was unnecessary. Second, ground-ground PTTs were counted by hand in the PVD Baseline and automatically by the VSCS in the DSR Baseline, as described in Section 5.4.3, which created an inconsistency between baselines in what was included as a PTT and what was excluded. We believe that the DSR Baseline values are valid and, if the VSCS and the same analysis techniques are used in future baselines, the data are suitable for other comparisons.

5.5 Usability

Figures 11a and b show ratings for several aspects of the DSR and PVD CHI. In general, the participants rated the PVD more favorably than the DSR. The only exception was for the radar and map displays where participants found the DSR equally easy to read and easier to understand than the PVD. Appendix C provides the participants' written responses to other items on the DSR keyboard and flight strip bays.

The review team concluded that these differences in usability ratings and comments resulted from "negative transfer" from the PVD. Negative transfer is a performance drop that occurs when highly automated skills on the old system are mistakenly used on the new system. For example, on the PVD, the QAKs that controllers press to begin data entries are located on the console, beside the main radar display. On the DSR, they are located on the top two rows of the keyboard. Because controllers use the QAKs continually, their use of these keys has become cognitively and physically automated. That is, they use the keys quickly and accurately but

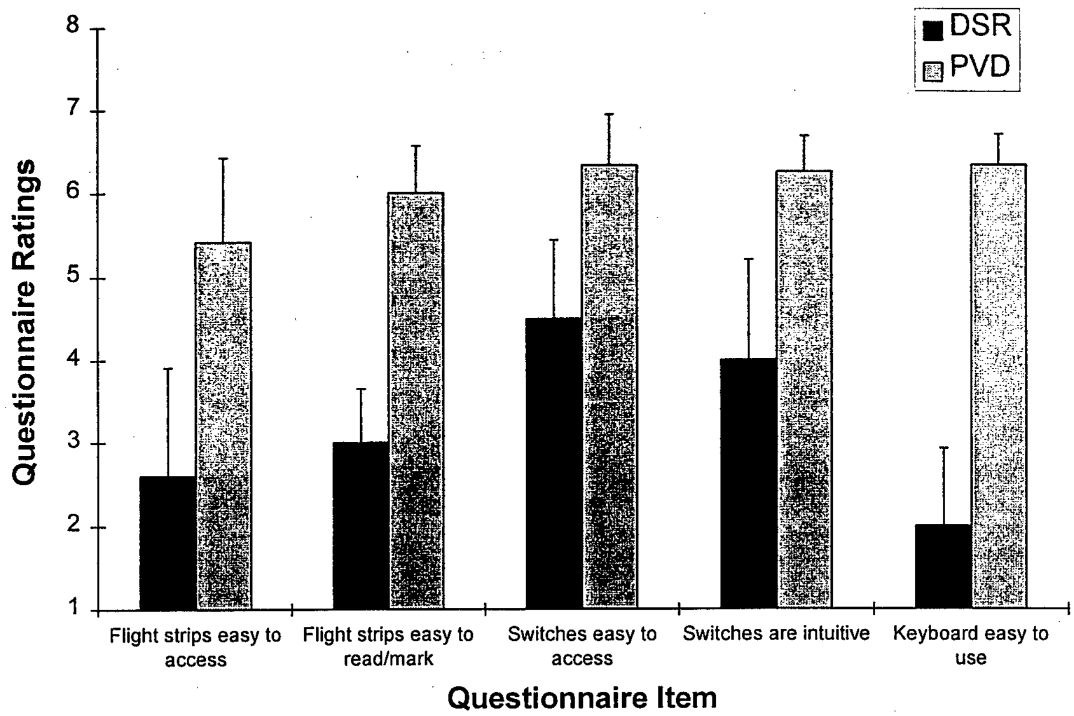

Figure 11a. Average ratings on the Final Questionnaire, Items A1-A5. Participants rated the extent to which they agreed with statements (1=strongly disagree, 8=strongly agree) about the DSR and PVD.

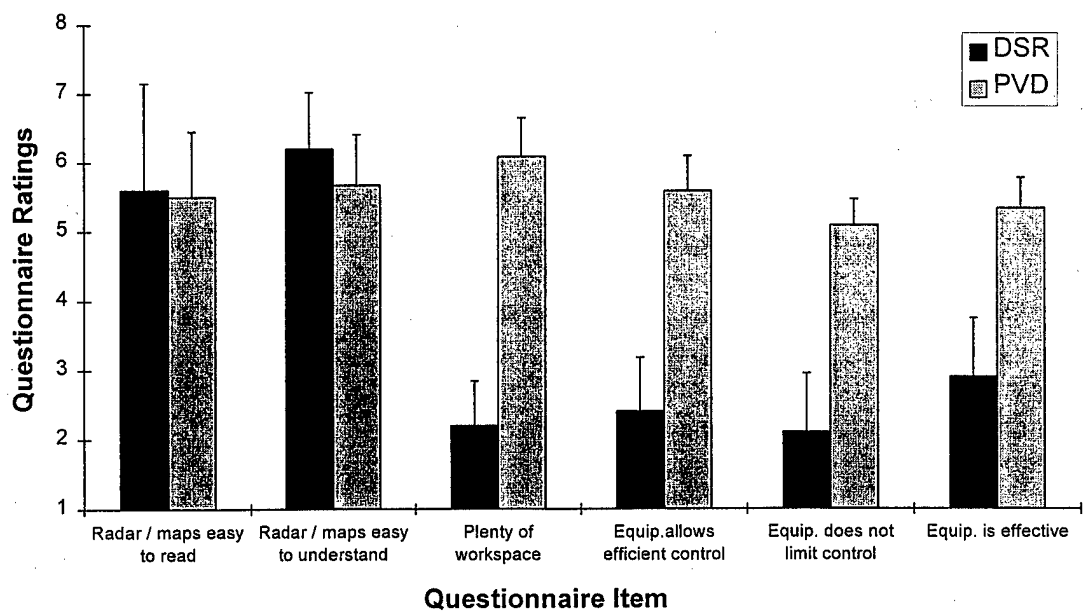

Figure 11b. Average ratings on the Final Questionnaire, Items A6-A11. Participants rated the extent to which they agreed with statements (1=strongly disagree, 8=strongly agree) about the DSR and PVD.

devote very little attention to the action. With the QAKs located on the DSR keyboard, however, their automatic skills were disrupted. Controllers automatically reached for the console instead of the keyboard or were forced to make entries more slowly or deliberately. Controllers were aware of their own performance. They recognized that they were making more errors and working more slowly. They may have interpreted this as resulting from poor design in the DSR rather than their inexperience using it. This probably resulted in their negative ratings about many aspects of the DSR.

It will take a great deal of experience (and not merely additional training) with the DSR keyboard to make the controllers' skills automatic again. It would be informative to re-administer the Final Questionnaire to the DSR Baseline participants after they have used the DSR in the field for several years. In that case, their skills would be automatic again, and the comparison to the PVD would be more valid.

5.6 Simulation Fidelity

5.6.1 Traffic Scenario Characteristics

The PVD Baseline and the DSR Baseline used the same traffic scenarios. However, because the baselines used different simulation platforms, the data analysis tools and techniques used to determine aircraft characteristics differed. This resulted in invalid comparisons for the arrival/departure/overflight and jet/propeller measures. Because of this, any differences found between baselines on these measures could be artifacts of the analysis method. We believe that the DSR Baseline values are valid and, if the TGF and the same analysis technique are used in future baselines, the data are suitable for other comparisons.

5.6.2 Realism Rating

As shown in Figures 12a and b, the participants rated the PVD Baseline scenarios as more realistic, especially the D controllers. The team concluded that this difference was mainly due to problems with the ghost sector and reduced D controller involvement in the DSR Baseline. As discussed earlier, a non-controller who accepted all handoffs and point outs staffed the ghost sector in the DSR Baseline. This reduced the D controllers' involvement with the scenario because they stopped doing between-sector coordination. In the field, coordination activities are one of the D controller's most important duties. Because the D controllers perceived that their duties were reduced, they rated the scenarios as unrealistic.

5.6.3 Impact of Technical Problems Rating

As shown in Figures 13a and b, the participants rated the DSR Baseline as having a smaller impact from technical problems than the PVD Baseline. In the DSR Baseline, 1 out of 13 runs was interrupted by technical problems. In the PVD Baseline, 1 out of 8 runs was interrupted. The team agreed that this contributed to a perception among the participants that the laboratory environment in the DSR Baseline was more stable than the PVD Baseline.

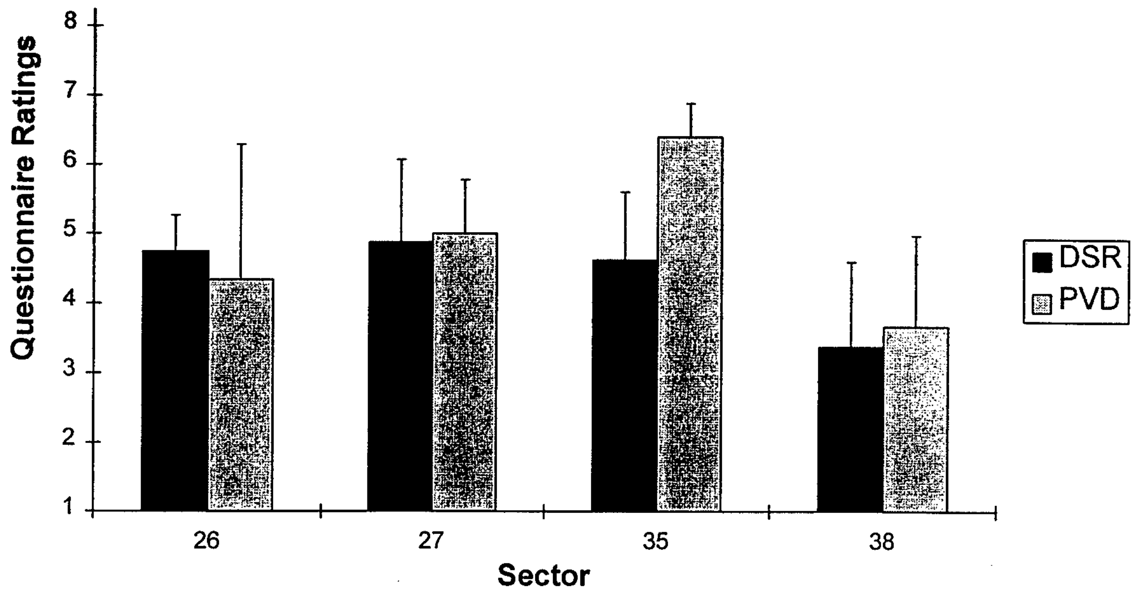

Figure 12a. Average ratings for the realism of the scenarios given by R controllers, averaged across runs. (1=Not Very Realistic, 8=Extremely Realistic).

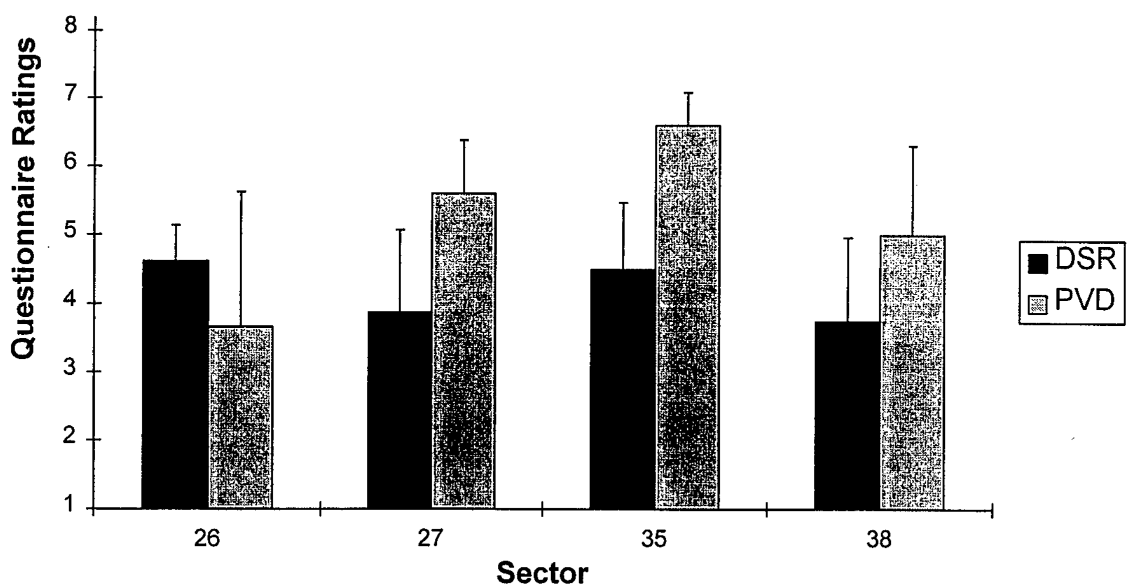

Figure 12b. Average ratings for the realism of the scenarios given by D controllers, averaged across runs. (1=Not Very Realistic, 8=Extremely Realistic).

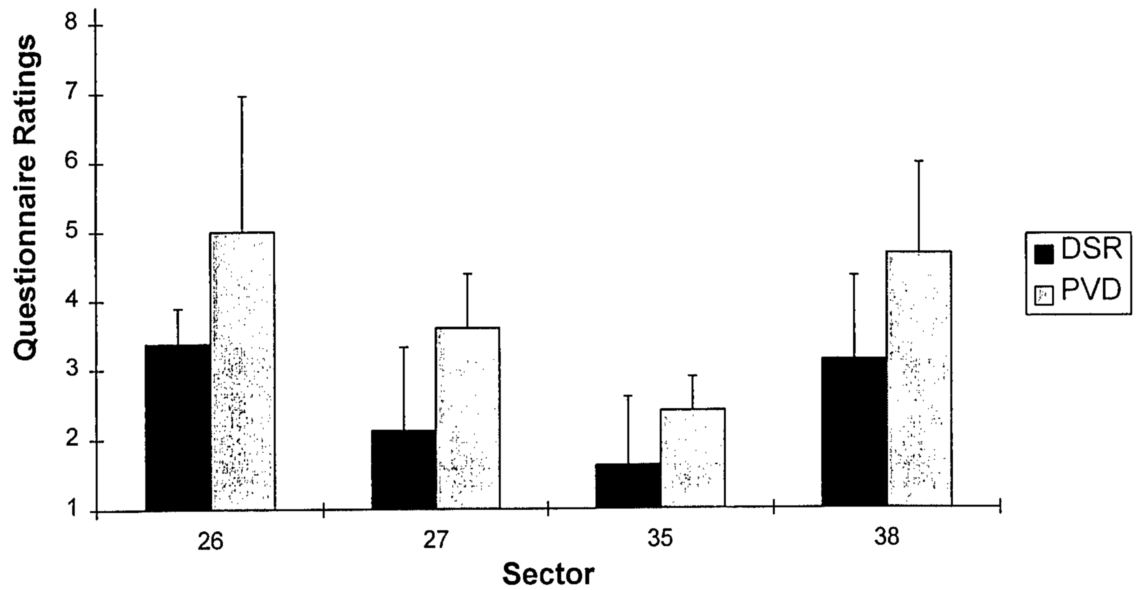

Figure 13a. Average ratings for the extent to which technical problems interfered with the participants' ability to control traffic, as given by R controllers, averaged across runs. (1=Not Much, 8=A Great Deal)

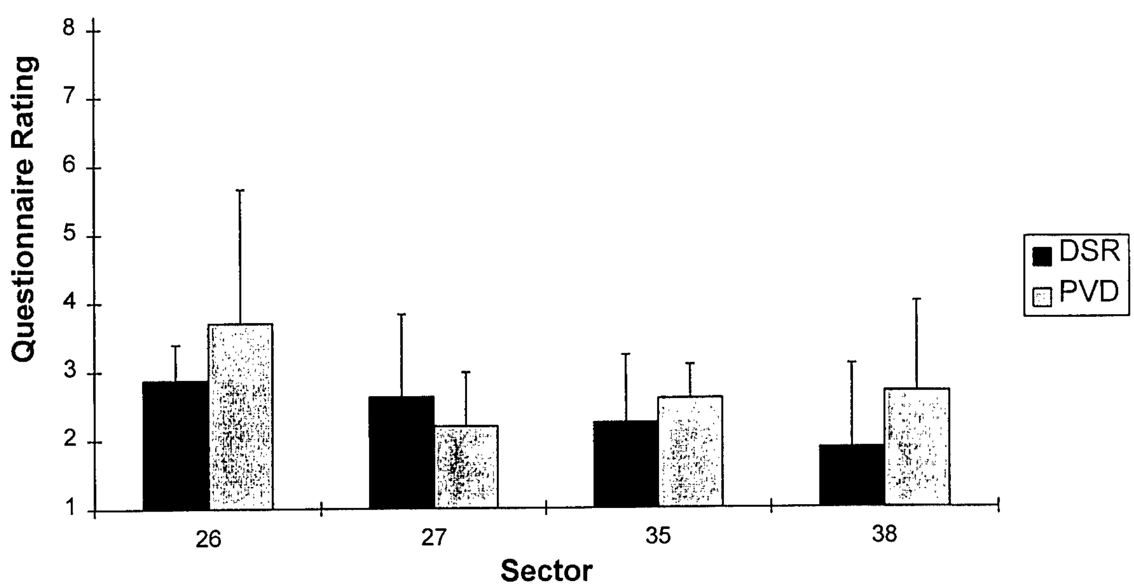

Figure 13b. Average ratings for the extent to which technical problems interfered with the participants' ability to control traffic, as given by D controllers, averaged across runs. (1=Not Much, 8=A Great Deal)

5.6.4 Impact of Pseudopilots' Rating

The Post-Scenario Questionnaire in the PVD Baseline did not include this questionnaire item, so no comparison is possible. However, the participants in the DSR Baseline gave low ratings on this item, indicating that problems associated with the pseudopilots were relatively minor and had little effect on their ability to control traffic.

5.6.5 Scenario Difficulty Rating

As shown in Figures 14a and b, the participants rated the scenarios as more difficult in the PVD Baseline than in the DSR Baseline. Because the two baselines used the same traffic scenarios, it is difficult to develop an explanation for this result. The team concluded that the reduced involvement of the D controllers and the different flight strip parameters contributed to this difference. In addition, traffic levels rise each year. It is possible that what was considered busy in 1992 when the data were originally collected was no longer considered busy by the controllers in 1997.

6. Recommendations for the DSR

6.1 Overlapping Data Blocks

Overlapping data blocks were more difficult to read in the DSR than in the PVD. The data show that the controllers using the DSR Baseline made more data block positioning entries than controllers using the PVD Baseline. This is a workload issue and could become a safety issue if critical information on the data blocks becomes obscured or too difficult to read. As traffic volume increases, more data blocks will overlap, and the readability problem will increase. We recommend that the Air Traffic DSR Evolution Team (ATDET) pursue several lines of action to address this problem.

First, graphical techniques known as anti-aliasing have been developed in recent years to make on-screen characters more readable, especially at small character sizes. These techniques use sophisticated manipulations of the character color to make curves appear smooth, to remove a pixeled appearance, and to create the illusion that the character is a continuous object. We recommend that the ATDET examine anti-aliasing techniques to determine if they are suitable for characters on radar displays and if they improve readability when data blocks overlap.

Second, in the past, the FAA has implemented functions to automatically reposition overlapping data blocks to improve data block readability. Each time, however, the controllers generally rejected these functions because the functions did not match how controllers actually use data blocks in the field. First, controllers do not like elements of their radar displays to change without an explicit action. When a data block moves automatically, it may distract the controller or unnecessarily draw attention from other information on the display. Second, controllers use data block positions to help them sort aircraft into categories and help them remember when

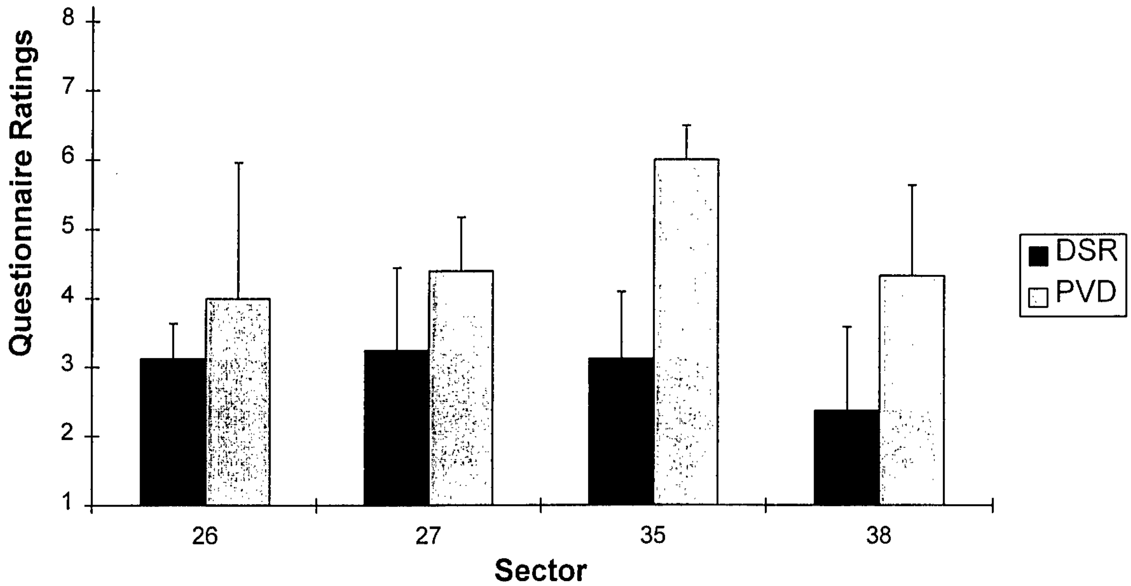

Figure 14a. Average ratings for the difficulty of the scenario, as given by R controllers, averaged across runs. (1=Not Very Difficult, 8=Extremely Difficult)

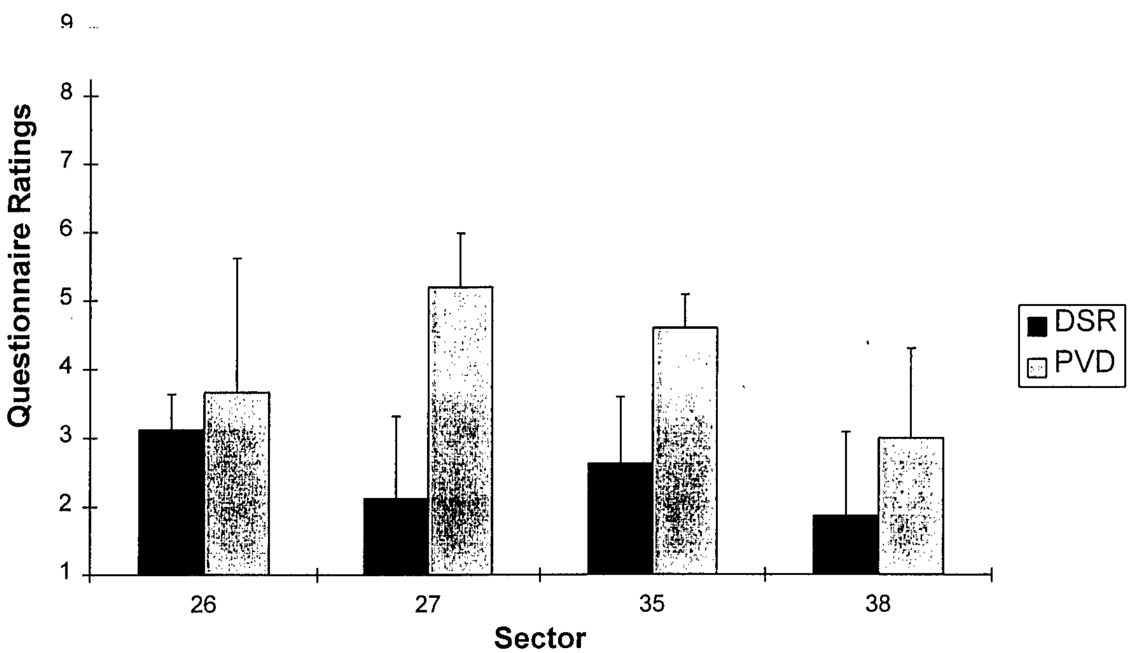

Figure 14b. Average ratings for the difficulty of the scenario, as given by D controllers, averaged across runs. (1=Not Very Difficult, 8=Extremely Difficult)

actions have been completed. For example, some controllers place the data blocks for all northbound traffic to the right of the target and all southbound traffic to the left of the target. This helps keep the data blocks separated but also serves as a memory aid as to where aircraft are headed. In another example, controllers can enter "/O" to indicate that they have told the aircraft to contact the next sector by changing the data block position and leader line length. This action reduces the leader line length to zero and positions the data block directly adjacent to the target. This reminds the controller that he or she has finished operations with that aircraft. Data block anti-overlap functions do not take this use of data blocks into account. As a result, a controller may position a data block to convey a specific meaning, but the anti-overlap function moves the data block and breaks the controller's memory aid.

We recommend that the ATDET and the appropriate Air Traffic Plans and Procedures (ATP) organizations develop new data block anti-overlap functions that meet the needs of controllers. We also recommend that some of the concepts developed by Eurocontrol for their ODID system be considered (Eurocontrol, 1999). In addition, we recommend that the ATDET group explore techniques other than data block position to code aircraft. The information contained in data blocks is crucial for safe ATC and should not be obscured. In the PVD, no other options for coding were available. The DSR, however, has many new capabilities, such as using different colors, that may be as good or better than data block position for coding aircraft. If techniques other than data block position can be used to categorize aircraft, anti-overlap functions may be more widely accepted by the controllers.

6.2 Flight Strip Bays

The DSR flight strip bays received much lower ratings than the PVD flight strip bays and many negative comments. The DSR bays were rated as difficult to access and as making the strips hard to read and mark. In particular, the participants' comments say that the strip bays were difficult to reach from the R position and that the sloped 22-strip bays made it difficult to write on the strips at the bottom of the bays.

We recommend that the DSR flight strip bays be examined from a formal ergonomics perspective. This would ensure that seated reach distance guidelines are met and would provide specific recommendations for improvements, if necessary. Guidelines for reach distances can be found in the FAA *Human Factors Design Guide for Acquisition of Commercial-of-the-Shelf, Non-Developmental Items, and Developmental Systems* (Wagner, Birt, Snyder, & Duncanson, 1996). We also recommend that the bays be examined against reach distance requirements of the *Americans with Disability Act Accessibility Guidelines* (Architectural and Transportation Barriers Compliance Board, 1994) to ensure that disabled personnel are able to complete their strip-related tasks effectively.

6.3 Keyboard

The DSR uses the same HCS software as the PVD and maintains the same data entry formats and syntax. However, the hardware used to make data entries in the two systems is substantially different and resulted in an increase in data entry errors in the DSR Baseline (see Section 5.3.3). This increase in data entry errors contributed to an overall increase in the number of data entries in the DSR Baseline because each incorrect entry must be re-entered correctly.

This increase in data entry errors was mostly due to participants' inexperience using the DSR keyboard. As their experience with the DSR keyboard grows, most of these data entry errors will disappear. Again, increased experience and not merely additional training will reduce the number of data entry errors. The controllers need more experience with the keyboard rather than more training. During the transition from PVD to DSR, then, we recommend that the number of D controllers be increased. These controllers can make many of the routine data entries and can offload data entry workload from the R controllers. In addition, the D position is keyboard intensive, which will give controllers more experience using the DSR keyboard.

In one case, however, we recommend that the ATDET re-evaluate the DSR keyboard design. Many controllers use PCs at home or in other non-ATC activities at the ARTCC. The backspace key is located in the upper right of every PC keyboard manufactured for the North American market. Even with experience using the DSR keyboard, controllers are still likely to press CLEAR when they intend to press BACKSPACE. Experience with the DSR keyboard will reduce this problem, but it is unlikely to completely eliminate it because of the pervasiveness of the traditional PC keyboard design.

Many of the participants mentioned the "sensitivity" of the DSR keyboard as a cause of many data entry errors. The sensitivity they mention is the result of the key travel (the distance a key must be moved downward before it is activated) and the key force (the amount of pressure that must be exerted on the key to activate it). We recommend that the ATDET conduct formal evaluations of the DSR keys to ensure that they follow the applicable ANSI standards (American National Standards Institute/Human Factors Society, 1988) for key travel and key force. If they do not, we recommend that future upgrades to the DSR incorporate changes to the keyboard to decrease the keyboard sensitivity and the resulting data entry errors.

6.4 Vector Lines

Vector lines show where the aircraft will be in the next 0, 1, 2, 4, or 8 min if the aircraft continues at the same speed and heading. Controllers typically work traffic with the vector lines set to 0-2 min. When issuing a new clearance, controllers often increase the vector-line length to 8 min as a quick way to ensure that it will not lead to separation violations in the near future. Using the PVD, controllers accomplish this by simply turning the knob all the way to the right, looking quickly at the radarscope, and then returning the knob to its original position. Using the DSR, however, controllers must move the cursor to the Display Controls and Status View and position the cursor over the VECTOR pick area. Then, the controllers must press the ENTER trackball several times to increase the length to 4 or 8 min and then press the PICK trackball button several times to decrease the length to its original value.

Because the PVD uses a mechanical knob, no quantitative data could be collected about how frequently or quickly controllers adjust the vector lines. However, based on the review team's feedback and the quantitative data that show increased use of the halo in the DSR Baseline, we believe that controllers in the DSR Baseline found the DSR vector-line function to be slower and more awkward than the PVD function. As a result, the participants reduced their use of the vector line and used the halo instead. However, the halo does not provide the same information as the vector lines. The halo shows a 5-nmi circle around the present position of an aircraft. It does not show where the aircraft will be in the future and does not allow the controller adjust the diameter of the circle. Future systems such as the Initial Conflict Probe will provide much better predictions of future aircraft positions than the vector lines. However, until those technologies are implemented, the vector lines will remain an important tool.

The ATDET has recognized the problem with the vector lines and has made a CHI change to address it since the DSR OT&E. If a controller presses the rightmost trackball button (HOME), the Display Controls and Status View (DCSV) will open and the cursor will be positioned over the VECTOR pick area automatically. This change eliminates the need for the controller to open the DCSV, locate the proper pick area, and then carefully position the cursor over it. However, the controller must still increase and decrease the length using multiple presses of the trackball buttons. This CHI change does improve the usability of the function but does not completely resolve the issue. We recommend that the DSR ATDET continue to monitor this issue and develop prototype CHI designs as necessary.

The DSR uses software controls for the display functions; therefore, collecting quantitative data about these functions is now viable. We recommend that baseline data be collected for each display function to guide future changes and improvements to the DSR CHI.

7. Recommendations for the Process

Besides reviewing the comparison data, the review team also developed several recommendations for future human-performance baseline comparisons. These recommendations attempt to address the consistency and validity problems encountered in this comparison. Many of these recommendations, especially those with broad applicability, were incorporated into the *Air Traffic Control System Baseline Methodology Guide* (Allendoerfer & Galushka, 1999).

7.1 Side-by-Side Comparisons of Systems

The 2½-year interim between the PVD and DSR Baselines created most of the problems. Numerous hardware, software, and personnel changes occurred in the Technical Center laboratories and simulation support facilities. In addition, new equipment was deployed to the field and new procedures and LOAs were developed. Finally, some of the PVD Baseline participants and observers were unavailable to participate in the DSR Baseline. These changes made it very difficult to preserve consistency between the baselines; as a result, validity suffered.

We recommend that future baseline comparisons collect data for both systems as part of a single, larger simulation activity. Participants would run the same traffic scenarios using both systems, alternating systems on each run or each day. This procedure would drastically reduce configuration management problems and would provide much tighter experimental control. It would also ensure that participants followed the same procedures for both systems. In addition, a side-by-side comparison would ensure that the controller and observer samples were the same for both systems.

A single side-by-side comparison would be costly in terms of financial, equipment, and labor resources. However, a side-by-side comparison would save time and money overall by reducing the need to organize, prepare, run, and analyze a separate simulation for each system. More important, a side-by-side comparison would ensure internal validity of the data and would dramatically improve the resulting comparison.

7.2 Precise Control of Testing Environment

Most tests and evaluations conducted in the Technical Center engineering laboratories do not require the level of experimental control required by psychological research. As a result, configuration management procedures at the laboratories and support facilities were not sufficiently detailed in several areas such as aircraft performance models. In addition, the requirements we provided to the technical personnel about experimental control were insufficient.

The review team recommends increased involvement by the Technical Center laboratory and simulation support personnel during the simulation planning stage. It is especially important to involve the personnel who will actually set up and configure the simulation hardware and software. Researchers managing the simulation must specify their requirements more precisely. It is crucial that researchers explain to technical personnel the high degree of experimental control that is needed and to provide specific guidance as to how that level of control can be obtained.

With development of laboratories like the Integration and Interoperability Facility (I^2F) that have actual field equipment but are more flexible and available for human factors research, it should become easier to schedule and coordinate studies with the necessary level of control.

7.3 Refinements to Data Reduction and Analysis Processes

In general, the DR&A for both baselines took too long and contained too many inconsistencies. Hardware, software, and personnel changes in the simulation support facilities prevented us from using some of the DR&A methods used in the PVD Baseline. As a result, we were forced to develop new DR&A methods for the DSR Baseline. Though we tried to ensure that the methods were equivalent, there is the possibility that the different methods introduced unknown biases into the data.

The review team recommends that standard definitions and algorithms be developed to compute each of the baseline metrics. Engineering research psychologists, personnel from the relevant Technical Center facilities, and SMEs from the field should develop these standards. This activity could result in a single DR&A tool that would compute the baseline metrics quickly and automatically. Such a tool would substantially reduce the time needed to analyze data and would ensure consistency and validity of analyses.

The review team strongly recommends that future comparisons continue the practice of reviewing data with SMEs from the field. In this baseline comparison, the review process was invaluable in that problems with the data were identified and that operationally meaningful explanations for results were written.

7.4 Additional Operational Input

Current controllers and supervisors are valuable and limited resources. Because many ZDC personnel were already participating in training and OT&E activities, long-term operational involvement was unavailable during the preparations for DSR Baseline. As a result, inconsistencies in procedures and LOAs were not identified until the data collection phase. Other inconsistencies, such as flight strip printing parameters, may also have been identified had earlier operational involvement been possible.

The review team recommends that current controllers and supervisors from the simulated facility participate throughout the baseline planning process. Their input is especially important during the shakedown and laboratory setup stages to ensure high levels of simulation realism. After shakedown, these individuals would be well suited to serve as SME Observers during data collection because of their experience with the simulation platform and scenarios.

More operational involvement in the planning stages will require a larger commitment of staff from the field. However, the benefit to the program would be substantial due to improved simulation realism and internal validity.

7.5 Increased Traffic Complexity

Though based on data from a 90th percentile day at ZDC, it is clear that the participants did not find the traffic scenarios used in the PVD and DSR Baselines as challenging as we intended. Not only did this prevent us from collecting data under high complexity conditions, it also reduced the motivation and interest of our participants.

We believe the 90th percentile day is still an appropriate traffic volume benchmark. However, we recommend closer examination of the recorded traffic data upon which scenarios are based. For example, the traffic volume metrics of the facility cover the entire ARTCC. Heavy volume for the whole facility does not necessarily mean heavy volume for an individual sector. If the traffic data are recorded in a relatively light sector, even on a busy day, the traffic scenarios developed from that data will not contain the necessary complexity. Psychologists and current SMEs from the field should thoroughly review traffic scenarios prior to shakedown to ensure that the traffic volume in the simulated sectors is as high as intended.

We also recommend developing a metric of traffic scenario complexity that is not tied exclusively to traffic volume. This metric would allow us to give a meaningful complexity score to the scenarios used in a simulation and would allow comparisons to scenarios in other studies. The dynamic density metric currently under development by the FAA should be suitable for this purpose.

7.6 Realistic Opportunities for Between-Sector Coordination

Between-sector coordination tasks constitute a substantial portion of a controller's job, especially at the D position. In the PVD Baseline, current ZDC controllers staffed the ghost sectors and ensured that participants followed applicable coordination procedures. In the DSR Baseline, however, participants who were unfamiliar with ZDC airspace and procedures staffed the ghost sectors. This enabled participants to request unrealistic point outs or to not coordinate at all. This probably contributed to the lower workload ratings made by participants during the DSR Baseline. It also reduced the internal validity of the baseline comparison and contributed to some data being discounted by the review team.

The review team recommends that current controllers from the field site staff all ghost sectors. Controllers who are not current, controllers from other facilities, and non-controllers do not have sufficient knowledge of the airspace and procedures to provide the required level of realism. Current controllers staffing the ghost sectors would be responsible for all communication and coordination over the frequency. They also would ensure that participants follow procedures and phraseology and would not accept handoffs or approve point outs that would normally be denied in the field. Staffing the ghost sector could easily be incorporated into the controller rotation schedule. Simulation support personnel could staff each ghost sector in addition to the controller to complete simulator-specific tasks such as deleting completed tracks.

Meeting this recommendation should restore D-controller workload to realistic levels for a given traffic volume. It will also help ensure that participants do not create unrealistic traffic situations by not following coordination procedures. Meeting this recommendation may require some additional staff from the simulated facility but should add considerable value by providing a much more realistic and consistent simulation.

7.7 Timing of Baseline Activities

We conducted the DSR Baseline during the second week of DSR OT&E. However, the length of time needed to reduce, analyze, and review the data from a baseline, assuming current DR&A tools and staffing, is several months. If the recommendations generated by the comparison were to guide CHI changes prior to deployment, we conducted the DSR Baseline too late in the acquisition process. The baseline must be conducted ahead of the first system deployment by a large enough margin that the recommendations generated from the baseline still can be incorporated into the system if necessary. If the DR&A procedures are improved as we recommend, we believe that the analysis, review, and reporting period can be reduced to around one month. Organizations responsible for setting a schedule should include time to conduct and analyze the baseline in their schedule and should also include time to address any issues generated during the baseline.

7.8 Scale of Baseline Activities

Some human factors issues may be better examined through small-scale, part-task evaluations rather than full-scale simulations. Full-scale simulations require that participants (rather than psychologists) decide when and how to take actions. This makes particular kinds of data such as speed and accuracy measurements very difficult to collect and analyze. On the other hand, in part-task evaluations, psychologists can specify when and how events occur and when particular actions are taken. They would allow psychologists to collect precise measurements under tightly controlled conditions.

Part task evaluations would be particularly useful to compare candidate design solutions earlier in the development process. For example, participants could complete a set of 50 flight plan entries using four different keyboard layouts. The speed and accuracy with which the participants completed the task could be measured. The order of presentation of the keyboards could be balanced and researchers could tightly control the exact characteristics of the flight plans.

The review team recognizes that not all studies that could have been conducted were possible given cost and schedule considerations. However, if additional human factors evaluations of the DSR are planned, the review team recommends several part-task evaluations to address specific issues raised with the DSR in this comparison. First, a part-task evaluation should be conducted to compare alternative DSR flight strip bay layouts. Participants could be given a series of strip-related actions to complete (e.g., locating, marking, and rearranging strips), and the speed and accuracy of these actions could be measured for both layouts. Second, a part-task evaluation should be conducted to compare alternative DSR keyboards. Participants could be given a series of data entries to complete, and their speed, accuracy, and heads-down time could be measured using the keyboards. Finally, a part-task evaluation should be conducted to compare alternative display controls to the DSR on-screen display controls. Participants could receive a set of display actions to complete, and the speed, accuracy, and heads-down time could be measured for both systems.

Finally, the review team recommends that future system comparisons use part-task evaluations early in the system evaluation process to examine specific human factors issues. The results of these evaluations would be provided to system vendors so that human factors improvements could be made prior to the human-performance baseline. In addition, the review team recommends that part-task evaluations continue after the baseline comparison to examine any remaining issues in detail.

8. Conclusion

This report identifies some of the inherent difficulties associated with medium-scale, high-fidelity, ATC simulation and controlled measurement of human performance and workload. Intervening variables stemming from the simulation platform and configuration management can confound results and limit the nature of conclusions that can be drawn. However, despite these limitations, we collected valuable objective data that may guide future system design, training, and procedural improvements for the DSR.

System baseline comparisons between FAA radar display systems had not been attempted before this effort, which increased the number of unknown factors. Fiscal and time constraints placed on the researchers also limited the planning and the execution of the baseline simulations. Despite these issues, system baselines are important to the future of FAA acquisitions and the system modernization process. Formalizing the role of human-performance baselines in the life cycle of FAA systems in conjunction with other human factors efforts will result in significant improvements in system development, evaluation, and operational use.

REFERENCES

Allendoerfer, K. R., & Galushka, J. (1999). *Air traffic control system baseline methodology guide* (DOT/FAA/CT-TN99/15). Atlantic City International Airport, NJ: Federal Aviation Administration William J. Hughes Technical Center.

American National Standards Institute (ANSI)/Human Factors Society (HFS). (1988). *American national standard for human factors engineering of visual display terminal workstations* (ANSI/HFS 100-1988). Santa Monica CA: Human Factors Society.

Architectural and Transportation Barriers Compliance Board. (1994). *Americans with Disabilities Act accessibility guidelines for buildings and facilities, transportation facilities, transportation vehicles* (NTIS No. PB95-256236INZ). Washington, DC: National Technical Information Service.

Eurocontrol. (1999, November 12). Label overlap prevention [Article posted on the World Wide Web]. Brussels, Belgium: Author. Retrieved December 21, 1999 from the World Wide Web: http://www.eurocontrol.be/projects/eatchip/hmi/show_room/acc/acc_overlap.html.

Federal Aviation Administration. (1996). Display System Replacement (DSR) overview. Washington, DC: Federal Aviation Administration En Route Domain Integration Team.

Galushka, J., Frederick, J., Mogford, R., & Krois, P. (1995). Plan View Display baseline research report (DOT/FAA/CT-TN95/45). Atlantic City International Airport, NJ: Federal Aviation Administration Technical Center.

Keegan, C., Skiles, T., Krois, P., & Merkle, M. (1996). Baseline measurement approach to ATC acquisitions. Journal of Air Traffic Control, 38, 33-37. Krois, P., & Marsden, A. (1997). FAA ODID VI: En route baseline comparison simulation final report (Report No. 311). Brétigny-sur-Orge Cedex, France: Eurocontrol Experimental Centre.

Krois, P., & Marsden, A. (1997). FAA ODID VI: En route baseline comparison simulation final report (Report No. 311). Brétigny-sur-Orge Cedex, France: Eurocontrol Experimental Centre.

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

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

Wagner, D., Birt, J. A., Snyder, M. D., & Duncanson, J. P. (1996). Human factors design guide: For the acquisition of commercial-off-the-shelf subsystems, non-developmental items, and developmental items (DOT/FAA/CT-96/1). Atlantic City International Airport, NJ: Federal Aviation Administration Technical Center.

Appendix A
DSR Baseline Questionnaires

BACKGROUND QUESTIONNAIRE

Controller: ___________ Date: _________
Team: ___________

Instructions

The purpose of this questionnaire is to obtain information concerning your experience and background. This information will be used to describe the participants in this study as a group. So that your identity can remain anonymous, your actual name should not be written on this form. Instead, your data will be identified by a controller code known only to yourself and the experimenters.

---

1) What is your age?
   _____ years

2) How many years have you actively controlled traffic?
   _____ years

3) How many hours have you used the DSR?
   _____ hours

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

5) What is your current position as an air traffic controller?
   ❑ Developmental ❑ Full Performance Level ❑ Other (specify) __________

6) In which environment do you have the most experience as an air traffic controller?
   ❑ En Route ❑ Terminal ❑ Other (specify) __________

7) If you wear corrective lenses, will you have them with you to wear during the simulation?
   ❑ Yes ❑ No ❑ I don't wear corrective lenses

8) Circle the number which best describes your current state of health.

1	2	3	4	5	6	7	8
Not Very Healthy							Extremely Healthy

9) Circle the number which best describes your current skill as an air traffic controller.

1	2	3	4	5	6	7	8
Not Very Skilled							Extremely Skilled

10) Circle the number which best describes your level of experience with personal computers.

1	2	3	4	5	6	7	8
Not Very Experienced							Extremely Experienced

11) Circle the number which best describes your level of satisfaction with the DSR.

1	2	3	4	5	6	7	8
Not Very Satisfied							Extremely Satisfied

POST-SCENARIO QUESTIONNAIRE

Controller: __________ Date: _________

Team: __________

Sector: 26 38 27 35 Run: 1 2 3 4

Position: Radar Data

Instructions

The purpose of this questionnaire is to obtain information concerning different aspects of the air traffic control problem just completed. This information will be used to determine how the simulation experience affects your opinions. As you answer each question, feel free to use the entire numerical scale. Please be as honest and as accurate as you can. So that your identity can remain anonymous, your actual name should not be written on this form. Instead, your data will be identified by a controller code known only to yourself and the experimenters.

---

1) How well did you control traffic during this problem?

1	2	3	4	5	6	7	8
Not Very Well							Extremely Well

2) What was your average workload level during this problem?

1	2	3	4	5	6	7	8
Very Low Workload							Very High Workload

3) How difficult was this problem compared to other simulation training problems?

1	2	3	4	5	6	7	8
Not Very Difficult							Extremely Difficult

4) How good do you think your air traffic control services were from a pilot's point of view?

1	2	3	4	5	6	7	8
Not Very Good							Extremely Good

5) To what extent did technical problems with the simulation equipment interfere with your ability to control traffic?

1	2	3	4	5	6	7	8
Not Very Much							A Great Deal

6) To what extent did problems with simulator pilots interfere with your normal air traffic control activities?

1	2	3	4	5	6	7	8
Not Very Much							A Great Deal

7) How realistic was this simulation problem compared to actual air traffic control?

1	2	3	4	5	6	7	8
Not Very Realistic							Extremely Realistic

OBSERVER EVALUATION FORM

Observer: __________ Date: __________
Controller: __________ Run: 1 2 3 4
Sector: 26 38 27 35
Position: Radar Data

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	Label 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 that 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

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

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

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

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

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 handoff 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

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

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

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

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

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

TECHNICAL KNOWLEDGE

18. Showing Knowledge of LOAs and SOPs

19. Showing Knowledge of Aircraft Capabilities and Limitations

20. Other Actions Observed in Technical Knowledge

COMMUNICATING

21. Using Proper Phraseology

22. Communicating Clearly and Efficiently

23. Listening to Pilot Readbacks and Requests

24. Other Actions Observed in Communicating

Observer Log

Observer: ____________ Date: ____________
Sector: 26 38 35 27 Run: 1 2 3 4

Instructions

Please record any unusual events by noting the system time, the nature of the event, and the aircraft involved. Please also note any technical problems and other safety-critical or otherwise important events. Use back of page for explanations, if necessary.

System Time	Event	Aircraft

FINAL QUESTIONNAIRE
Controller: __________ Date: _________
Team: __________

Section A

Please circle the number which best describes your level of agreement with each of the following statements concerning the DSR.

1) The flight progress strips are easy to access in the strip bays.

1	2	3	4	5	6	7	8
Strongly Disagree							Strongly Agree

2) The flight progress strips are easy to read and mark in the strip bays.

1	2	3	4	5	6	7	8
Strongly Disagree							Strongly Agree

3) The on-screen controls are easy to access.

1	2	3	4	5	6	7	8
Strongly Disagree							Strongly Agree

4) The operation and functions of the on-screen controls are intuitive.

1	2	3	4	5	6	7	8
Strongly Disagree							Strongly Agree

5) The controller keyboard is easy to use.

1	2	3	4	5	6	7	8
Strongly Disagree							Strongly Agree

6) The radar and map displays are easy to read.

1	2	3	4	5	6	7	8
Strongly Disagree							Strongly Agree

7) The radar and map displays are easy to understand.

1	2	3	4	5	6	7	8
Strongly Disagree							Strongly Agree

8) There is plenty of space to work within the workstation.

1	2	3	4	5	6	7	8
Strongly Disagree							Strongly Agree

9) The equipment, displays, and controls allow me to control traffic in the most efficient way possible.

1	2	3	4	5	6	7	8
Strongly Disagree							Strongly Agree

10) The equipment, displays, and controls allow me to control traffic without any awkward limitations.

1	2	3	4	5	6	7	8
Strongly Disagree							Strongly Agree

11) Overall, the equipment, displays and controls are effective in meeting the needs of controllers.

1	2	3	4	5	6	7	8
Strongly Disagree							Strongly Agree

Section B

Please circle the number which best describes your overall interaction with the equipment, displays, and controls (i.e., human-computer interface) of the DSR.

1)	1 Not Very Limited	2	3	4	5	6	7	8 Extremely Limited
2)	1 Not Very Frustrating	2	3	4	5	6	7	8 Extremely Frustrating
3)	1 Not Very Effective	2	3	4	5	6	7	8 Extremely Effective
4)	1 Not Very Efficient	2	3	4	5	6	7	8 Extremely Efficient
5)	1 Not Very Easy to Operate	2	3	4	5	6	7	8 Extremely Easy to Operate
6)	1 Not Very Easy to Understand	2	3	4	5	6	7	8 Extremely Easy to Understand

Section C

Please circle the number which best represents your opinion about the following potential improvements to the DSR system.

1) To what extent do you think a light-colored map display (e.g., tan) with dark letters would improve your effectiveness with the DSR console?

1	2	3	4	5	6	7	8
Not Very Much							A Great Deal

2) To what extent do you think a mouse input device (instead of a trackball) would improve your effectiveness with the DSR console?

❑ If you are not familiar with a mouse input device, mark this box.

1	2	3	4	5	6	7	8
Not Very Much							A Great Deal

3) To what extent do you think additional color-coding of information would improve your effectiveness with the DSR?

1	2	3	4	5	6	7	8
Not Very Much							A Great Deal

4) To what extent do you think a brighter room lighting level would improve your effectiveness with the DSR console?

1	2	3	4	5	6	7	8
Not Very Much							A Great Deal

Section D

For each the following questions. indicate your opinion by marking one or more of the provided boxes. Then, please provide any additional comments that you think are appropriate.

1) Which aspects of the DSR console need improvement?

- ❑ Radar and Map Displays
- ❑ Flight Strip Bays
- ❑ Keyboard
- ❑ Trackball
- ❑ On-Screen Controls
- ❑ Volume of Workspace
- ❑ Other (specify) ________________
- ❑ Other (specify) ________________

Please provide some details about why you think each of these aspects needs improvement?

____
____
____
____
____
____
____
____

2) What are the most common mistakes you encounter using the DSR console?

- ❑ Misreading Radar Display Information
- ❑ Misreading Map Display Information
- ❑ Misreading Flight Progress Strips
- ❑ Making Entries with Keyboard
- ❑ Selecting Targets with Trackball
- ❑ Adjusting On-screen Controls
- ❑ Other (specify) ________________
- ❑ Other (specify) ________________

Please provide some details about what you think causes you to make each of these mistakes?

____
____
____
____
____
____
____

Section E

If there are any other comments or suggestions that you have regarding this baseline study of the DSR console, please write your ideas in the space provided below.

Air Traffic Workload Input Technique Rating

Controller: __________ Date: ___________

You have been using the seven-key ATWIT system to rate your workload during the Baseline Study. Please indicate below how you define the lowest (1) and highest (7) workload rating on the seven-point ATWIT scale.

To me, the lowest ATWIT rating (1) means my workload is:

To me, the highest ATWIT rating (7) means my workload is:

DSR Keyboard Questionnaire

Controller: __________ Date: __________

Team: __________ DSR hours this

week: __________

Instructions

The PVD/M1 and DSR keyboards differ in a number of ways. During DSR development and Operational Test and Evaluation (OT&E) activities, concerns have been raised about several properties of the DSR keyboard. Each concern is listed in the left column below. Please indicate whether or not you experienced that concern this week. If you answer "Yes" for a particular concern, then complete the four items in the right column. These items are:

- **How many times did this occur during the week?** Circle your estimate of how often you made this keyboard mistake this week while working with the DSR. Please estimate the frequency only for you and not for controllers in general. We realize that this is only a "best guess," but it will help us understand what you perceive to be the most frequent problems.
- **To what extent did this impact your efficiency?** Estimate how much this problem reduced your ability to control air traffic efficiently. Please estimate the impact on only your efficiency and not on the ability of controllers in general. We realize that this is only a "best guess," but it will help us understand what you perceive to be the most serious problems.
- **How did you correct it?** If you took an action to correct the mistake, please describe what you did. For example, "I backspaced over it and typed the correct letter." Please describe only what actions you took and not the actions other controllers took or could take.
- **To what extent will this improve with experience?** Estimate the extent to which this problem will occur less frequently as you gain experience with the DSR keyboard. Please estimate only your own rate of improvement and not the rate of improvement for controllers in general. We realize that this is only a "best guess," but it will help us understand what you perceive to be the most persistent problems.

When I type entries, I look primarily at (check one):

- ☐ Message Composition Area
- ☐ Keyboard/Hands
- ☐ Situation Display
- ☐ Other (please specify): __________

Concern	If "Yes"
1. The BACKSPACE and CLEAR keys are in different locations on the DSR keyboard than on the PVD/M1 keyboard. This week, did you make any mistakes that you could attribute to the location of these keys? ❑ Yes ❑ No	How many times did this occur during the week? 1-5 times 5-10 times 10 or more times To what extent did this impact your efficiency? 1 2 3 4 5 6 7 Very Little ... A Great Deal How did you correct it? To what extent will this improve with experience? 1 2 3 4 5 6 7 Very Little ... A Great Deal
2. Some function keys on the DSR have different labels than the equivalent QAKs on the PVD/M1. This week, did you make any mistakes that you could attribute to these labels? ❑ Yes ❑ No	How many times did this occur during the week? 1-5 times 5-10 times 10 or more times To what extent did this impact your efficiency? 1 2 3 4 5 6 7 Very Little ... A Great Deal How did you correct it? To what extent will this improve with experience? 1 2 3 4 5 6 7 Very Little ... A Great Deal
3. Some function keys on the DSR keyboard are grouped differently than on the PVD/M1. This week, did you make any mistakes that you could attribute to these groupings? ❑ Yes ❑ No	How many times did this occur during the week? 1-5 times 5-10 times 10 or more times To what extent did this impact your efficiency? 1 2 3 4 5 6 7 Very Little ... A Great Deal How did you correct it? To what extent will this improve with experience? 1 2 3 4 5 6 7 Very Little ... A Great Deal

Concern	If "Yes"
4. Function keys are located on the DSR keyboard rather than on the PVD/M1 console. This week, did you make any mistakes that you could attribute to this function key placement? ❑ Yes ❑ No	How many times did this occur during the week? 1-5 times 5-10 times 10 or more times To what extent did this impact your efficiency? 1 2 3 4 5 6 7 Very Little … A Great Deal How did you correct it? To what extent will this improve with experience? 1 2 3 4 5 6 7 Very Little … A Great Deal
5. The keys on the DSR keyboard are easier to press than the keys on the PVD/M1. This week, did you make any mistakes that you could attribute to this difference in key sensitivity? ❑ Yes ❑ No	How many times did this occur during the week? 1-5 times 5-10 times 10 or more times To what extent did this impact your efficiency? 1 2 3 4 5 6 7 Very Little … A Great Deal How did you correct it? To what extent will this improve with experience? 1 2 3 4 5 6 7 Very Little … A Great Deal
6. The MAP MAN and MAP PSET keys are located on the top row of the DSR keyboard. This week, did you made any mistakes that you could attribute to the location of these keys? ❑ Yes ❑ No	How many times did this occur during the week? 1-5 times 5-10 times 10 or more times To what extent did this impact your efficiency? 1 2 3 4 5 6 7 Very Little … A Great Deal How did you correct it? To what extent will this improve with experience? 1 2 3 4 5 6 7 Very Little … A Great Deal

Concern	If "Yes"
7. Some controllers have reported inadvertently hitting the SPACE key on the numeric keypad. Did you experience this during the week? ❑ Yes ❑ No	How many times did this occur during the week? 1-5 times 5-10 times 10 or more times To what extent did this impact your efficiency? 1 2 3 4 5 6 7 Very Little ... A Great Deal How did you correct it? To what extent will this improve with experience? 1 2 3 4 5 6 7 Very Little ... A Great Deal
8. Some controllers have reported that the 0 key on the numeric keypad is not easily identifiable to touch users. Did you experience this during the week? ❑ Yes ❑ No	How many times did this occur during the week? 1-5 times 5-10 times 10 or more times To what extent did this impact your efficiency? 1 2 3 4 5 6 7 Very Little ... A Great Deal How did you correct it? To what extent will this improve with experience? 1 2 3 4 5 6 7 Very Little ... A Great Deal
9. Some controllers have suggested that the 0 key and the SPACE key on the numeric keypad should be swapped. This week, did you ever make any mistakes that you could attribute to the location of these two keys? ❑ Yes ❑ No	How many times did this occur during the week? 1-5 times 5-10 times 10 or more times To what extent did this impact your efficiency? 1 2 3 4 5 6 7 Very Little ... A Great Deal How did you correct it? To what extent will this improve with experience? 1 2 3 4 5 6 7 Very Little ... A Great Deal

Concern	If "Yes"
10. Some controllers have reported inadvertently hitting the INSRT or DEL CHAR keys while trying to hit the 0 key on the numeric keypad. Did you experience this during the week? ❑ Yes ❑ No	How many times did this occur during the week? 1-5 times 5-10 times 10 or more times To what extent did this impact your efficiency? 1 2 3 4 5 6 7 Very Little ... A Great Deal How did you correct it? To what extent will this improve with experience? 1 2 3 4 5 6 7 Very Little ... A Great Deal
11. Some controllers have suggested that the numeric keypad keys are too close together. This week, did you make any mistakes that you could attribute to the closeness of these keys? ❑ Yes ❑ No	How many times did this occur during the week? 1-5 times 5-10 times 10 or more times To what extent did this impact your efficiency? 1 2 3 4 5 6 7 Very Little ... A Great Deal How did you correct it? To what extent will this improve with experience? 1 2 3 4 5 6 7 Very Little ... A Great Deal
12. Some controllers have reported being distracted by the tone that sounds after hitting the blank key on the numeric keypad. Did you experience this during the week? ❑ Yes ❑ No	How many times did this occur during the week? 1-5 times 5-10 times 10 or more times To what extent did this impact your efficiency? 1 2 3 4 5 6 7 Very Little ... A Great Deal How did you correct it? To what extent will this improve with experience? 1 2 3 4 5 6 7 Very Little ... A Great Deal

Concern	If "Yes"
13. Some controllers have reported inadvertently pressing the HOME button on the trackball when they meant to press the ENTER button. Did you experience this during the week? ❑ Yes ❑ No	How many times did this occur during the week? 1-5 times 5-10 times 10 or more times To what extent did this impact your efficiency? 1 2 3 4 5 6 7 Very Little ... A Great Deal How did you correct it? To what extent will this improve with experience? 1 2 3 4 5 6 7 Very Little ... A Great Deal

Appendix B
DSR Baseline Measurement Summary

Table B-1. Overall Level Data

Construct	Measure	Average	Standard Deviation	Comments
Safety	Operational Errors	0.0	N/A	
	Conflict Alerts	1.7	2.19	
	Halo Initiations	4.9	3.52	
	Data Block Positioning	93.7	40.34	
	Other Safety Critical Issues	N/A	N/A	Handwritten data. See Section 5.1.5.
Capacity	Aircraft Under Control	50.6	12.58	
	Time in Sector	N/A	N/A	See Table B-2 for Sector Level Data.
Performance	Overall Data Entries-R	309.5	74.47	
	Overall Data Entries-D	46.8	48.00	
	Specific Data Entry Types	N/A	N/A	See Table B-3 for Sector Level Data for specific data entry types.
	Data Entry Errors-R	34.0	16.40	

Construct	Measure	Average	Standard Deviation	Comments
Performance (continued)	Data Entry Errors-D	9.1	10.19	
	Number of Altitude, Speed, and Heading Changes	50.6	24.52	
	Quality of ATC services from a controller point of view-R	7.4	1.12	
	Quality of ATC services from a controller point of view-D	7.3	1.44	
	Quality of ATC services from a pilot point of view-R	7.3	0.99	
	Quality of ATC services from a pilot point of view-D	6.8	1.56	
	Maintaining Safe and Efficient Traffic Flow-R	6.3	1.14	
	Maintaining Safe and Efficient Traffic Flow-D	6.5	1.00	
	Maintaining Attention and Situation Awareness-R	6.1	1.06	
	Maintaining Attention and Situation Awareness-D	5.5	1.17	
	Prioritizing-R	6.0	1.02	

Construct	Measure	Average	Standard Deviation	Comments
Performance (continued)	Prioritizing-D	5.9	1.11	
	Providing Control Information-R	5.3	1.80	
	Providing Control Information-D	6.9	0.60	
	Technical Knowledge-R	6.2	1.12	
	Technical Knowledge-D	6.2	1.20	
	Communicating-R	5.7	1.37	
	Communicating-D	5.7	1.10	
Workload	ATWIT Workload-R	2.6	0.74	
	ATWIT Workload-D	2.3	0.66	
	Post-Run Workload-R	3.3	1.07	
	Post-Run Workload-D	2.8	1.09	
	Communication Taskload (Number of Air-Ground Communications)	139.8	28.59	
	Coordination Taskload (Number of Ground-Ground Communications)	21.3	25.48	

Construct	Measure	Average	Standard Deviation	Comments
Usability	Flight Progress Strip Access	2.6	2.12	
	Flight Progress Strip Read/Mark	3.0	1.05	
	Ease of Access to Controls	4.5	1.51	
	Operation of Controls Intuitive	4.0	1.94	
	Keyboard Ease of Use	2.0	1.49	
	Radar and Map Ease of Use	5.6	2.50	
	Radar and Map Ease of Understanding	6.2	1.32	
	Workstation Space	2.2	1.03	
	Equipment, Displays, and Controls Support Efficient ATC	2.4	1.26	
	Equipment, Displays, and Controls Impose Limitations	2.1	1.37	
	Equipment, Displays, and Controls Overall Effectiveness	2.9	1.37	

Construct	Measure	Average	Standard Deviation	Comments
Simulation Fidelity	Traffic Scenario Characteristics			It is not meaningful to average these data across traffic scenarios. Please see Table B-2 for Sector Level Data.
	Scenario Duration	N/A	N/A	
	Number of Jet Aircraft	N/A	N/A	
	Number of Propeller Aircraft	N/A	N/A	
	Number of Arrivals	N/A	N/A	
	Number of Departures	N/A	N/A	
	Number of Overflights	N/A	N/A	
	Realism Rating-R	4.4	1.64	See Table B-2 for Sector Level Data.
	Realism Rating-D	4.2	1.54	
	Impact of Technical Problems Rating-R	2.5	2.05	
	Impact of Technical Problems Rating-D	2.4	1.48	
	Impact of Pseudopilots Rating-R	1.8	0.75	

Construct	Measure	Average	Standard Deviation	Comments
Simulation Fidelity (continued)	Impact of Pseudopilots Rating-D	1.9	0.89	
	Scenario Difficulty Rating-R	2.9	1.20	
	Scenario Difficulty Rating-D	2.5	0.96	

Table B-2. Sector Level Data - Averages

Construct	Measure	26	27	35	38	Comments
Safety	Operational Errors	0.0	0.0	0.0	0.0	
	Conflict Alerts	1.4	1.9	2.4	1.1	
	Halo Initiations	6.2	8.0	3.3	2.0	
	Data Block Positioning	76.0	111.0	123.6	64.0	
Capacity	Aircraft Under Control	39.3	58.5	66.4	38.1	
	Time in Sector	8.8	7.9	12.6	7.4	
Performance	Overall Data Entries-R	306.6	363.9	328.4	239.1	
	Overall Data Entries-D	33.6	51.8	55.6	46.3	
	Data Entry Errors-R	26.4	34.9	42.3	32.6	
	Data Entry Errors-D	4.3	8.6	10.0	13.4	
	Number of Altitude, Speed, and Heading Changes	65.5	76.0	29.8	31.3	
	Quality of ATC services from a controller point of view-R	7.4	7.5	7.5	7.3	

Construct	Measure	26	27	35	38	Comments
Performance (continued)	Quality of ATC services from a controller point of view-D	7.6	6.9	7.4	7.3	
	Quality of ATC services from a pilot point of view-R	7.1	7.1	7.3	8.0	
	Quality of ATC services from a pilot point of view-D	7.3	6.3	6.9	6.6	
	Maintaining Safe and Efficient Traffic Flow-R	6.5	6.3	6.0	6.5	
	Maintaining Safe and Efficient Traffic Flow-D	6.7	6.3	7.0	6.0	
	Maintaining Attention and Situation Awareness-R	6.7	5.8	6.3	6.0	
	Maintaining Attention and Situation Awareness-D	5.4	5.9	5.6	4.8	
	Prioritizing-R	6.5	6.1	5.5	5.8	
	Prioritizing-D	5.8	6.3	5.9	5.7	
	Providing Control Information-R	6.0	5.5	3.9	6.0	
	Providing Control Information-D	7.0	7.0	7.0	6.0	
	Technical Knowledge-R	6.2	6.1	6.1	6.7	

Construct	Measure	26	27	35	38	Comments
Performance (continued)	Technical Knowledge-D	6.4	6.6	6.1	5.7	
	Communicating-R	6.3	5.6	5.3	5.8	
	Communicating-D	6.3	5.6	6.0	5.0	
Workload	ATWIT Workload-R	2.8	2.6	2.8	1.9	
	ATWIT Workload-D	2.4	2.1	2.7	1.8	
	Post-Run Workload-R	3.4	3.3	3.5	3.0	
	Post-Run Workload-D	3.0	2.9	3.4	1.9	
	Communication Taskload (Number of Air-Ground Communications)	132.3	154.8	163.6	103.1	
	Coordination Taskload (Number of Ground-Ground Communications)	17.0	17.8	28.3	22.1	
Simulation Fidelity	Traffic Scenario Characteristics					
	Scenario Duration	60 minutes (70 min - 10 min ramp)	90 minutes (100 min - 10 min ramp)	90 minutes (100 min - 10 min ramp)	60 minutes (70 min - 10 min ramp)	

Construct	Measure	26	27	35	38	Comments
Simulation Fidelity (continued)	Number of Jet Aircraft	20.9	35.4	62.4	38.1	
	Number of Propeller Aircraft	18.6	22.6	4.3	0.4	
	Number of Arrivals	11.0	39.4	9.0	10.1	
	Number of Departures	26.4	13.6	41.0	26.4	
	Number of Overflights	2.0	5.0	16.7	2.0	
	Realism Rating-R	4.8	4.9	4.6	3.4	
	Realism Rating-D	4.6	3.9	4.5	3.8	
	Impact of Technical Problems Rating-R	3.4	2.1	1.6	3.1	
	Impact of Technical Problems Rating-D	2.9	2.6	2.3	1.9	
	Impact of Pseudopilots Rating-R	2.0	1.6	1.6	1.8	
	Impact of Pseudopilots Rating-D	2.4	2.1	1.5	1.8	
	Scenario Difficulty Rating-R	3.1	3.3	3.1	2.4	
	Scenario Difficulty Rating-D	3.1	2.1	2.6	1.9	

Table B-3. Sector Level Data – Standard Deviations

Construct	Measure	26	27	35	38	Comments
Safety	Operational Errors	NA	NA	NA	NA	
	Conflict Alerts	0.79	2.30	3.54	1.33	
	Halo Initiations	3.31	3.11	2.59	1.51	
	Data Block Positioning	32.31	31.75	46.98	14.76	
Capacity	Aircraft Under Control	2.60	1.41	1.19	3.09	
	Time in Sector	0.89	0.78	1.19	1.20	
Performance	Overall Data Entries-R	76.75	62.46	49.45	53.52	
	Overall Data Entries-D	46.27	58.57	49.17	43.55	
	Data Entry Errors-R	17.70	15.06	15.21	16.51	
	Data Entry Errors-D	4.46	12.51	8.60	12.65	
	Number of Altitude, Speed, and Heading Changes	17.72	15.46	12.08	6.94	
	Quality of ATC services from a controller point of view-R	0.92	1.07	0.76	1.75	

Construct	Measure	26	27	35	38	Comments
Performance (continued)	Quality of ATC services from a controller point of view-D	0.74	1.73	1.41	1.75	
	Quality of ATC services from a pilot point of view-R	1.13	0.99	1.16	0.00	
	Quality of ATC services from a pilot point of view-D	0.89	1.58	1.36	2.20	
	Maintaining Safe and Efficient Traffic Flow-R	0.84	1.75	1.00	0.55	
	Maintaining Safe and Efficient Traffic Flow-D	1.53	0.96	0.00	1.41	
	Maintaining Attention and Situation Awareness-R	0.82	1.39	0.49	1.26	
	Maintaining Attention and Situation Awareness-D	0.89	0.90	1.30	1.47	
	Prioritizing-R	0.55	1.27	0.93	0.98	
	Prioritizing-D	0.98	1.11	1.25	1.21	
	Providing Control Information-R	1.17	1.05	1.46	0.82	
	Providing Control Information-D	1.26	1.51	2.42	0.00	
	Technical Knowledge-R	1.17	1.05	1.46	0.82	

Construct	Measure	26	27	35	38	Comments
Performance (continued)	Technical Knowledge-D	1.34	0.53	1.55	1.21	
	Communicating-R	0.82	1.33	1.83	1.17	
	Communicating-D	0.52	0.98	0.93	1.55	
Workload	ATWIT Workload-R	0.85	0.52	0.62	0.65	
	ATWIT Workload-D	0.57	0.66	0.46	0.64	
	Post-Run Workload-R	1.19	0.71	1.60	0.76	
	Post-Run Workload-D	0.93	0.99	1.19	0.64	
	Communication Taskload (Number of Air-Ground Communications)	13.85	19.30	20.94	11.08	
	Coordination Taskload (Number of Ground-Ground Communications)	16.46	23.87	33.25	28.95	
Simulation Fidelity	Traffic Scenario Characteristics					
	Scenario Duration	NA	NA	NA	NA	
	Number of Jet Aircraft	1.46	1.27	0.98	3.76	
	Number of Propeller Aircraft	1.13	0.79	0.76	0.79	

Construct	Measure	26	27	35	38	Comments
Simulation Fidelity (continued)	Number of Arrivals	0.58	0.79	0.82	1.46	
	Number of Departures	2.15	1.27	1.15	3.15	
	Number of Overflights	0.00	0.00	0.49	0.00	
	Realism Rating-R	0.71	1.25	1.77	2.26	
	Realism Rating-D	1.41	1.64	1.31	1.83	
	Impact of Technical Problems Rating-R	2.50	0.99	0.52	3.04	
	Impact of Technical Problems Rating-D	1.64	2.26	0.71	0.64	
	Impact of Pseudopilots Rating-R	0.93	0.74	0.74	0.71	
	Impact of Pseudopilots Rating-D	0.92	1.13	0.53	0.71	
	Scenario Difficulty Rating-R	1.46	0.71	1.55	0.92	
	Scenario Difficulty Rating-D	0.83	0.64	1.06	0.83	

Table B-4. Sector Level Data – Specific Data Entry Types Averages

Data Entry Type	26-R	26-D	27-R	27-D	35-R	35-D	38-R	38-D
AM	0.0	2.1	0.0	1.9	0.0	2.6	0.0	2.8
CO	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
FP	0.0	0.0	0.0	0.0	0.0	1.1	0.0	0.0
FR	0.0	2.3	0.0	2.6	0.0	4.9	0.0	2.0
LA	1.4	0.0	3.8	0.0	2.4	0.0	1.1	0.0
LB	0.9	0.0	0.3	0.0	1.3	0.0	0.8	0.0
SG	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
QA	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0
QB	1.3	0.3	1.4	0.0	0.5	0.8	0.1	0.0
QD	1.4	0.4	1.6	0.3	1.0	0.1	1.6	1.1
QF	3.5	0.0	2.9	0.0	1.6	0.0	2.5	0.0
QN	116.4	1.0	175.4	1.6	178.0	0.6	105.5	2.1
QP	14.4	6.7	18.8	3.5	7.5	12.9	8.8	26.6
QQ	41.0	3.6	48.6	1.6	7.0	0.8	22.6	3.4
QT	1.3	0.1	2.9	11.8	0.4	0.4	0.5	0.1
QU	17.8	5.3	13.0	2.0	5.5	5.0	5.8	2.8
QX	0.1	0.0	0.1	0.1	0.1	0.8	0.0	0.4
QZ	5.3	19.0	10.6	26.6	3.1	25.5	3.6	19.8
SR	0.0	1.9	0.0	1.9	0.0	6.8	0.0	1.0

Table B-5. Sector Level Data – Specific Data Entry Types Standard Deviations

Data Entry Type	26-R	26-D	27-R	27-D	35-R	35-D	38-R	38-D
AM	0.00	1.73	0.00	3.59	0.00	2.23	0.00	2.99
CO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FP	0.00	0.00	0.00	0.00	0.00	2.62	0.00	0.00
FR	0.00	1.67	0.00	2.23	0.00	7.20	0.00	2.40
LA	1.11	0.00	1.98	0.00	3.08	0.00	1.05	0.00
LB	0.93	0.00	0.66	0.00	1.48	0.00	1.09	0.00
SG	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00
QA	0.00	0.00	0.33	0.00	0.00	0.00	0.33	0.00
QB	1.39	0.45	1.11	0.00	0.71	0.83	0.33	0.00
QD	0.70	0.73	0.70	0.66	0.71	0.33	0.99	2.62
QF	3.43	0.00	2.26	0.00	1.32	0.00	1.94	0.00
QN	30.33	1.60	34.03	1.32	42.01	0.86	28.84	1.90
QP	5.94	9.27	6.61	3.32	6.65	20.55	4.66	51.49
QQ	11.21	6.46	21.20	2.60	2.92	1.09	6.00	5.02
QT	1.20	0.35	3.55	27.46	0.70	0.99	1.00	0.33
QU	9.04	8.17	11.77	2.24	6.84	5.77	7.45	3.19
QX	0.33	0.00	0.33	0.33	0.33	1.39	0.00	0.99
QZ	3.19	28.98	10.99	21.38	3.33	31.92	2.96	21.73
SR	0.00	1.96	0.00	2.57	0.00	9.31	0.00	1.50

Table B-6. Interval Level Data – Sector 26 - Averages

Construct	Measure	1	2	3	4	5
Safety	Data Block Positioning	6.0	16.1	28.0	19.0	18.6
Capacity	Aircraft Under Control	9.3	16.8	15.0	12.3	12.5
Performance	Overall Data Entries-R	23.6	57.1	96.0	67.5	62.4
	Overall Data Entries-D	3.5	5.0	12.6	8.8	3.8
	Data Entry Errors-R	3.8	4.9	5.9	6.3	6.4
	Data Entry Errors-D	0.8	1.1	1.6	0.5	0.3
	Number of Altitude, Speed, and Heading Changes	3.0	20.9	17.5	9.8	14.0
Workload	ATWIT Workload-R	1.8	3.0	2.9	2.9	3.1
	ATWIT Workload-D	1.6	2.6	2.7	2.4	2.3
	Communication Taskload (Number of Air-Ground Communications)	14.7	34.7	34.5	23.6	24.2

Table B-7. Interval Level Data – Sector 27 - Averages

Construct	Measure	1	2	3	4	5	6	7
Safety	Data Block Positioning	7.5	17.6	21.4	9.9	15.9	21.0	20.3
Capacity	Aircraft Under Control	8.9	14.6	11.8	4.9	9.3	17.3	14.3
Performance	Overall Data Entries-R	31.4	59.0	58.4	25.0	41.9	79.4	68.9
	Overall Data Entries-D	2.1	7.6	8.4	1.5	6.9	10.5	14.8
	Data Entry Errors-R	2.4	4.3	5.6	1.9	5.9	8.5	6.4
	Data Entry Errors-D	0.1	1.5	1.4	0.3	2.5	1.9	1.0
	Number of Altitude, Speed, and Heading Changes	5.1	9.6	15.1	3.8	4.8	18.0	19.6
Workload	ATWIT Workload-R	1.9	2.7	2.9	1.9	2.2	3.5	3.3
	ATWIT Workload-D	1.5	2.3	2.0	1.4	2.0	2.9	2.5
	Communication Taskload (Number of Air-Ground Communications)	17.2	32.3	31.1	10.9	18.8	44.6	37.9

Table B-8. Interval Level Data – Sector 35 - Averages

Construct	Measure	1	2	3	4	5	6	7
Safety	Data Block Positioning	8.9	16.0	13.8	19.8	29.3	23.0	13.0
Capacity	Aircraft Under Control	9.1	15.1	17.6	23.0	25.0	20.3	13.4
Performance	Overall Data Entries-R	23.8	36.5	37.9	57.3	69.6	58.9	44.5
	Overall Data Entries-D	13.1	7.1	5.5	7.9	8.8	7.5	5.8
	Data Entry Errors-R	4.1	4.1	4.0	8.0	6.8	7.9	7.4
	Data Entry Errors-D	0.8	1.1	1.0	2.1	1.5	2.5	1.0
	Number of Altitude, Speed, and Heading Changes	1.1	3.5	5.6	6.4	5.6	5.1	2.4
Workload	ATWIT Workload-R	1.8	2.5	2.7	3.5	4.1	3.3	2.1
	ATWIT Workload-D	1.6	2.5	2.9	3.5	3.4	2.8	2.4
	Communication Taskload (Number of Air-Ground Communications)	12.2	23.8	25.0	34.8	39.2	28.8	19.4

Table B-9. Interval Level Data – Sector 38 - Averages

Construct	Measure	1	2	3	4	5
Safety	Data Block Positioning	10.0	15.1	25.7	14.3	13.6
Capacity	Aircraft Under Control	11.4	12.6	13.3	7.9	10.0
Performance	Overall Data Entries-R	32.6	46.4	66.8	45.5	47.9
	Overall Data Entries-D	5.5	8.8	15.0	9.3	7.8
	Data Entry Errors-R	6.0	5.5	8.1	8.9	4.1
	Data Entry Errors-D	1.4	3.5	4.5	2.1	1.9
	Number of Altitude, Speed, and Heading Changes	5.9	6.6	10.0	4.1	4.5
Workload	ATWIT Workload-R	1.7	2.2	2.5	1.7	1.9
	ATWIT Workload-D	1.1	1.4	2.2	1.8	1.7
	Communication Taskload (Number of Air-Ground Communications)	17.7	20.1	31.4	14.6	16.7

Table B-10. Interval Level Data – Sector 26 – Standard Deviations

Construct	Measure	1	2	3	4	5
Safety	Data Block Positioning	4.62	6.96	13.52	14.57	12.11
Capacity	Aircraft Under Control	3.37	1.16	1.41	1.16	2.78
Performance	Overall Data Entries-R	13.42	18.29	49.55	16.89	28.89
	Overall Data Entries-D	4.11	4.87	20.64	18.50	4.03
	Data Entry Errors-R	5.26	4.26	3.94	8.07	6.29
	Data Entry Errors-D	1.04	1.13	2.26	0.76	0.46
	Number of Altitude, Speed, and Heading Changes	3.07	9.08	3.96	2.25	6.21
Workload	ATWIT Workload-R	0.57	0.68	1.08	1.10	1.07
	ATWIT Workload-D	0.49	0.65	0.62	0.40	0.61
	Communication Taskload (Number of Air-Ground Communications)	2.65	6.61	3.55	3.48	9.40

Table B-11. Interval Level Data – Sector 27 - Standard Deviations

Construct	Measure	1	2	3	4	5	6	7
Safety	Data Block Positioning	2.45	5.85	11.37	3.87	8.95	6.23	7.05
Capacity	Aircraft Under Control	0.83	0.92	1.39	1.55	1.28	1.98	1.28
Performance	Overall Data Entries-R	12.05	7.07	12.35	9.04	23.27	10.82	6.96
	Overall Data Entries-D	2.75	8.02	13.66	2.83	8.22	8.75	25.50
	Data Entry Errors-R	2.33	3.41	4.21	1.46	6.45	4.57	4.98
	Data Entry Errors-D	0.35	2.33	2.72	0.46	4.44	2.42	1.60
	Number of Altitude, Speed, and Heading Changes	1.13	2.77	4.02	2.60	2.82	5.50	5.93
Workload	ATWIT Workload-R	0.43	0.73	0.56	0.64	0.65	0.94	0.84
	ATWIT Workload-D	0.53	0.77	0.73	0.43	0.69	1.32	0.91
	Communication Taskload (Number of Air-Ground Communications)	4.01	5.66	5.03	2.67	8.25	4.36	10.01

Table B-12. Interval Level Data – Sector 35 - Standard Deviations

Construct	Measure	1	2	3	4	5	6	7
Safety	Data Block Positioning	4.26	8.11	5.50	9.48	8.31	12.28	7.07
Capacity	Aircraft Under Control	1.46	1.46	0.92	1.07	2.62	1.16	1.19
Performance	Overall Data Entries-R	8.05	10.27	6.24	13.47	8.19	10.72	12.46
	Overall Data Entries-D	20.50	9.03	7.19	7.77	9.91	7.48	6.36
	Data Entry Errors-R	2.53	1.96	1.20	3.12	1.83	2.90	9.18
	Data Entry Errors-D	1.16	2.42	1.07	2.70	1.93	2.62	0.76
	Number of Altitude, Speed, and Heading Changes	1.13	2.00	4.03	3.02	4.50	3.40	2.13
Workload	ATWIT Workload-R	0.73	0.97	1.05	0.87	0.58	0.75	0.71
	ATWIT Workload-D	0.68	0.62	0.56	0.66	0.66	0.64	0.51
	Communication Taskload (Number of Air-Ground Communications)	1.70	4.29	4.28	4.92	8.34	5.48	5.14

Table B-13. Interval Level Data – Sector 38 - Standard Deviations

Construct	Measure	1	2	3	4	5
Safety	Data Block Positioning	9.29	3.08	7.80	5.77	7.09
Capacity	Aircraft Under Control	4.21	2.67	1.16	0.99	2.73
Performance	Overall Data Entries-R	18.80	9.21	11.96	18.97	26.13
	Overall Data Entries-D	5.55	7.59	15.49	5.73	16.73
	Data Entry Errors-R	7.75	2.33	4.45	11.89	3.72
	Data Entry Errors-D	1.60	4.14	4.34	1.81	4.12
	Number of Altitude, Speed, and Heading Changes	2.64	2.07	4.14	1.46	2.39
Workload	ATWIT Workload-R	0.58	0.66	1.01	0.54	0.66
	ATWIT Workload-D	0.25	0.74	0.99	0.47	0.52
	Communication Taskload (Number of Air-Ground Communications)	1.81	1.75	5.37	3.28	6.63

Appendix C
Controller Comments

The following data represent controllers' partially edited responses to the Final Questionnaire, sections D and E. Responses are organized by controller and by each section of the questionnaire.

CONTROLLER 1

Section D
1. Displays
Flight Strip Bays
Keyboard
Comments: Flight strip bays inaccessible from R side. Keyboard keys are too closely grouped and far too sensitive.
2. Misreading Flight Progress Strips
Making Entries with Keyboard
Comments: Keyboard keys are too closely grouped and far too sensitive. Numerous re-entries.

Section E
Comments: System close to usable but keyboard and flight strip bays must be redesigned.

CONTROLLER 2

Section D
1. Displays
Flight Strip Bays
Keyboard
On-screen Controls
Workspace
Comments: Strips are difficult to see and reach. The design needs to be more user friendly. Keyboard is not intuitive. The key positions are awkward. The numeric keys need more space between them and ENTER button should be within easy reach of numeric keypad to allow one hand operation without having to look at the keyboard. Radar/Map Display and on-screen controls are hard to locate. Also, there is not enough variation in display brightness. I suggest color variations for on-screen controls. Volume of workspace is cramped and very limited.
2. Misreading Map Display Information
Making Entries with Keyboard
Adjusting On-screen Controls
Comments: See question 1 comments. Misreading map display-seeing on screen controls.

Section E
Comments: The simulations were very slow and not a true test of the system. I would have liked to see a more challenging simulation to test the system.

CONTROLLER 3

Section D

1. Flight Strip Bays
Keyboard
On-screen Controls
Workspace

Comments: The bays are too far away and too high to reach. I found it easier to stop marking tickets. [For] bays 2, 3, 4, and 5, I would have to get out of my chair to mark the tickets. The number pad is not workable. The keys are too close and the Space/Insert/Delete keys need to be taken out. CRD needs more options for set-up. For example, take out the code list once the sector has been set-up. Not enough [work]space for pad of paper.

2. Making Entries with Keyboard
Selecting Targets with Trackball
Adjusting On-screen Controls

Comments: The keyboard causes me the most problems. The way it is designed forces me to look at it for every entry, which distracts me from the radar. The trackball selecting seems to be to picky. Finding the [on-screen control] is too time consuming, taking my attention away from the radar.

Section E

blank

CONTROLLER 4

Section D

1. Flight Strip Bays
Workspace
Other: R functions from D side.

Comments: Flight strip bays inaccessible from R side. Not enough space for writing material and not enough space with tracker plugged in. D-side functions very hard to perform from R side when in one-person configuration.

2. Making Entries with Keyboard

Comments: Location of "?" and space keys and sensitivity of keyboard.

Section E

blank

CONTROLLER 5

Section D

1. Displays

Flight Strip Bays

Keyboard

On-screen Controls

Comments: Radar/map displays are too grainy. Map lines need sharper contrast. Flight strip bays-curved bottom on 22 bay a little difficult to use. Keyboard- Numeric keypad difficult to locate 8/0, can't operate blind (eyes off), fingers glance off intended key and cause double entries of intended characters and extraneous characters. On-screen controls—should be able to hide radar and strobe 1-4 (not utilized). CRD should be opaque.

2. Making Entries with Keyboard

Other: Trackball pick/enter

Comments: Keyboard-hitting clear button when wishing to use backspace key, hitting two keys (numeric) instead of one key. Trackball pick/enter keys- trying to remember which key to use when utilizing trackball re-route, track stunt, and range bearing functions.

Section E

Comments: Overhead map displays - The two Plexiglas sheets to be manually compressed to read center portion of chart, otherwise it is blurry. Strip holders for used strips - if strip is dropped behind VSCS VMD, may drop within VDM box - possible creating a fire hazard.

CONTROLLER 6

Section D

1. Flight Strip Bays

Keyboard

Workspace

Other: Strip Bay Lighting

Other: Location of map light and strip light controls

Comments: Strip bay lighting-too much shadowing (fluorescent is better). Flight strip bays-curved bay is awkward, tickets tend to fall out during sequencing. All lighting controls should be within easy reach of a seated R controller. Keyboard- keys too sensitive and number keys need to be spread out and isolated. Workspace-not enough room on the D side. Also could be more on R.

2. Making Entries with Keyboard

Comments: Trying to hit "slant 0"; hitting the zero key without looking or slowing down.

Section E

blank

CONTROLLER 7

Section D

1. Flight Strip Bays
Keyboard
Workspace
Other: Light controls
Comments: Flight strip bays are difficult to reach and have no support for writing. I get confused with some keys on keyboard - "0" "Backspace" "clear" and their positions of the keyboard. There needs to be more workspace at the D position. Lighting controls are difficult to reach.
2. Making Entries with Keyboard
Comments: It is difficult to make most entries without looking at the keyboard. The key placement is very confusing.

Section E

blank

CONTROLLER 8

Section D

1. Keyboard
Trackball
Workspace
Comments: The keyboard is too sensitive. I find myself making multiple entries due to incorrect inputs, or slow and deliberate inputs to ensure acceptance. The workspace provided is the minimum of what is absolutely necessary. More workspace would be more comfortable.
2. Making Entries with Keyboard
Comments: Some unfamiliarity with the keyboard but not much. I had approximately 40+ hours on it. Keys are too close too sensitive and too small.

Section E

blank

CONTROLLER 9

Section D
1. Flight Strip Bays
Keyboard
Workspace
Comments: Strip bay makes marking tickets difficult, no hand support, and strips are too small. Keyboard number pad difficult to work with. Keys too close together. Insert key & Delete key not needed. Consoles are too cramped.
2. Making Entries with Keyboard
Adjusting On-screen Controls
Comments: Keyboard number pad errors. The brightness controls seem vague and confusing. The master brightness affects the other brightness controls too much.

Section E
blank

CONTROLLER 10
Section D
1. Displays
Flight Strip Bays
Keyboard
Trackball
On-screen Controls
Workspace
Other: Strip Bay Lighting
Comments: Radar and map displays- Maps are hard to read: spiral lines of airways are hard to tell from map boundaries. Flight Strip Bays- Too much re-sequencing. Too far to reach from D side. You spend a lot of time with hands above shoulders. Keyboard is not user friendly. Fingers slide off keys and hit other keys. Trackball- cord is too short and too many buttons. On-screen controls- There is too much to look for if [you are] trying to find one brightness control, etc. [With the PVD] everything is in a separate place [in the DSR it is] all together.
2. Making Entries with Keyboard
Selecting Targets with Trackball
Adjusting On-screen Controls
Comments: Keyboard is just a poor design. It does not appear that any thought went into this at all. Keys are not grouped in any logical fashion. Number pad is awful. This keyboard matches no other keyboard I have ever worked on.

Section E
I don't think these problems were nearly as busy as the original PVD Baseline - also the data collection system was changed from PVD baseline to DSR baseline. (1-4 busy to 1-7 busy).