

2

USAFSAM-TP-90-26

AD-A234 626

AIRCREW EVALUATION SUSTAINED OPERATIONS PERFORMANCE (AESOP): A Triservice Facility for Technology Transition

**Samuel G. Schiflett, Ph.D. (USAFSAM/VNB)
David R. Strome, Ph.D. (Systems Research Labs., Inc.)
Douglas R. Eddy, Ph.D. (NTI, Inc.)
Mathieu A. Dalrymple, B.S. (Systems Research Labs., Inc.)**

**Systems Research Laboratories, Inc.
A Division of Arvin/Calspan
2800 Indian Ripple Road
Dayton, OH 45440-3696**

**NTI, Inc.
4130 Linden Avenue, Suite 235
Dayton, OH 45432**

December 1990

Interim Report for Period September 1987 - September 1990

Approved for public release; distribution is unlimited.



**Prepared for
USAF SCHOOL OF AEROSPACE MEDICINE
Human Systems Division (AFSC)
Brooks Air Force Base, TX 78235-5301**

**Sponsored by
Headquarters
US Army Medical R&D Command
Fort Detrick, MD 21701-5012**



DTIC FILE COPY

NOTICES

This interim paper was submitted by Systems Research Laboratories, Inc., A Division of Arvin/Calspan, 2800 Indian Ripple Road, Dayton, Ohio, and NTI, Incorporated, Dayton, Ohio, under contract F33615-87-D-0601, job order 7930-19-10, with the USAF School of Aerospace Medicine, Human Systems Division, AFSC, Brooks Air Force Base, Texas. Dr. Samuel Schiflett (USAFSAM/VNBC) was the Laboratory Project Scientist-in-Charge. Partial funding was received from the US Army Medical Research & Development Command, Ft. Detrick, Maryland.

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency, contractor, or subcontractor thereof. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency, contractor, or subcontractor thereof.

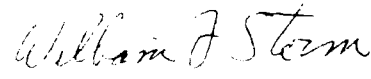
When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The voluntary informed consent of the subjects used in this research was obtained in accordance with AFR 169-6.

The Office of Public Affairs has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.


SAMUEL G. SCHIFLETT, Ph.D.
Project Scientist


WILLIAM F. STORM, Ph.D.
Supervisor


GEORGE E. SCHWENDER, Colonel, USAF, MC. CFS
Commander

REPORT DOCUMENTATION PAGE				Form Approved OMB No 0704-0188	
1a REPORT SECURITY CLASSIFICATION Unclassified			1b RESTRICTIVE MARKINGS		
2a SECURITY CLASSIFICATION AUTHORITY			3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.		
2b DECLASSIFICATION/DOWNGRADING SCHEDULE					
4 PERFORMING ORGANIZATION REPORT NUMBER(S)			5. MONITORING ORGANIZATION REPORT NUMBER(S) USAFSAM-TP-90-26		
6a. NAME OF PERFORMING ORGANIZATION Systems Research Laboratories NTI		6b OFFICE SYMBOL (if applicable)	7a. NAME OF MONITORING ORGANIZATION USAF School of Aerospace Medicine (VNB)		
6c. ADDRESS (City, State, and ZIP Code) (SRL) Dayton, OH 45440-3696 (NTI) Dayton, OH 45432			7b. ADDRESS (City, State, and ZIP Code) Human Systems Division (AFSC) Brooks AFB, Texas 78235-5301		
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U.S. Army Medical R&D Command		8b OFFICE SYMBOL (if applicable) SGRD/PLE	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F33615-87-D-0601		
8c. ADDRESS (City, State, and ZIP Code) HQ USAMRDC Ft. Detrick, MD 21701-5012			10 SOURCE OF FUNDING NUMBERS		
			PROGRAM ELEMENT NO. 63764A	PROJECT NO. 7930	TASK NO. 19
			WORK UNIT ACCESSION NO 10		
11. TITLE (Include Security Classification) Aircrew Evaluation Sustained Operations Performance (AESOP): A Triservice Facility for Technology Transition					
12. PERSONAL AUTHOR(S) Schiflett, Samuel (USAFSAM/VNB); Strome, David R.*; Eddy, Douglas R.**; Dalrymple, Mathieu*					
13a. TYPE OF REPORT Interim		13b. TIME COVERED FROM 87/09 TO 90/09		14. DATE OF REPORT (Year, Month, Day) 1990, December	
				15. PAGE COUNT 33	
16 SUPPLEMENTARY NOTATION * Systems Research Laboratories, Inc. ** NTI, Incorporated					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB-GROUP	AWACS Team Assessment Team Measurement		
05	08		Methodology for teams Team Effectiveness Team Performance		
23	02		Military teams Team Functions Team Processes		
19 ABSTRACT (Continue on reverse if necessary and identify by block number) A primary mission of the Sustained Operations Branch, Crew Technology Division, Armstrong Laboratory (AFSC), is to develop procedures and provide guidance to operational commands on maintaining and extending crew performance during sustained operations and continuous duty. The Crew Technology Division developed the Aircrew Evaluation Sustained Operations Performance (AESOP) facility under the sponsorship of the Office of Military Performance Assessment Technology to meet the triservice research and mission requirements for team performance metrics. The AESOP facility evaluates the interactive effects of fatigue, stress, and medications on Airborne Warning and Control Systems (AWACS) aircrew performance so effective counter measures can be transitioned from the laboratory to field test environments to actual operations. The basic design of the facility allows the flexibility and experimental control to either assess performance decrements or develop performance enhancement techniques in a realistic operational environment. The simulation integrates hardware and software resources, data collection and analysis systems, verbal communication networks, command and control scenarios, and performances measures. Examples of individual and team performance on complex decision making are illustrated. Emphasis is placed on flexibility of measurement, hierarchical organization of measurement levels, and data collection from multiple perspectives. Future research opportunities for the development and evaluation of candidate models to describe and predict team decision making under stress from the perspective of operational performance-based criteria are identified. (Partially supported by Army Medical Research and Development Command.)					
20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION Unclassified		
22a NAME OF RESPONSIBLE INDIVIDUAL Samuel G. Schiflett, PhD			22b TELEPHONE (Include Area Code) (512) 536-3464		22c OFFICE SYMBOL USAFSAM/VNB

PREFACE

A primary mission of the Sustained Operations Branch, Crew Technology Division, USAF School of Aerospace Medicine (USAFSAM), is to develop procedures and provide guidance to operational commands on maintaining and extending crew performance during sustained operations and continuous duty.

The USAFSAM developed the Aircrew Evaluation Sustained Operations Performance (AESOP) facility under the sponsorship of the Office of Military Performance Assessment Technology (OMPAT), formerly the Chemical Defense Joint Working Group on Drug Dependent Degradation of Military Performance (JWGD³ MILPERF), to meet the triservice research and mission requirements for team performance metrics. Continuous technical guidance was received from OMPAT during the development of the AESOP facility. Dr. Frederick Hegge, OMPAT's director, was especially helpful.

The AESOP facility, originally developed at the USAFSAM but recently transferred to the Armstrong Laboratory¹ at Brooks Air Force Base, is currently evaluating the interactive effects of fatigue, stress, and medications on Airborne Warning and Control Systems (AWACS) aircrew performance so effective countermeasures can be transitioned from the laboratory to field test environments to actual operations. The basic design of the facility allows the flexibility and experimental control to either assess performance decrements or develop performance enhancement techniques in a realistic operational environment. The simulation integrates hardware and software resources, data collection and analysis systems, verbal communication networks, command and control scenarios, and performance measures. Examples of individual and team performance on complex decision making are illustrated. Emphasis is placed on flexibility of measurement, hierarchical organization of measurement levels, and data collection from multiple perspectives. Future research opportunities for the development and evaluation of candidate models to describe and predict team decision making under stress from the perspective of operational performance-based criteria are identified. (Partially supported by Army Medical Research and Development Command.)

Approved for	
By	☒
Date	
Initials	
Justified	
By	
Distribution	
Availability	
Dist	Copy to
A-1	

¹ Sustained Operations Branch, Crew Technology Division, Detachment 4, Armstrong Laboratory, Brooks Air Force Base, TX 78235-5000.

TABLE OF CONTENTS

	Page
Primary Mission	1
Transition of Knowledge	1
Triservice-Supported Facility	1
A Unique Command and Control Research Facility	2
AESOP as a Complex-Decision Research Environment	2
AESOP - A Multipurpose Research Tool	4
Applications	8
Performance-Based Measures of Individual and Team Complex Decision Making ..	8
The Task	10
Performance Measurement	10
Hierarchy of Performance Measures	10
Problems in Measuring Performance in a Complex, Two-sided Environment	15
Methods	15
Additional Measures	15
Development of Realistic and Complex Command, Control, and Communications Scenarios	18
C ³ Environment Model	18
Command, Control, and Communications Generic Workstation (C ³ GW)	18
Developing the C ³ GW	18
Scenario Development	20
Task Loading	21
Embedded Tasks	21
Standardization of Design	23
Scenario Differentiation	23
Scenario Refinement	23
Running the Simulation	24
State-of-the-Art Research	24
Future Performance Research	25
Recommendation	25
References	26

LIST OF FIGURES

Figure Number		Page
1.	AESOP facility floor plan	5
2.	Cluster and simulation network	6
3.	Data recording devices	7
4.	AESOP associate networks	9
5.	Performance measurement hierarchy	12
6.	Levels of performance measurement	13
7.	C ³ antihistamine study	16
8.	C ³ research design	16
9.	C ³ environment model	19
10.	Weapons Director task loading events	22

AIRCREW EVALUATION SUSTAINED OPERATIONS PERFORMANCE (AESOP)

A Triservice Facility for Technology Transition

Primary Mission

A primary mission of the Sustained Operations Branch, Crew Technology Division, USAF School of Aerospace Medicine, Brooks Air Force Base, Texas, is to develop procedures and provide guidance to USAF operational commands on how to maintain and extend crew performance during sustained and continuous operations. Long-range deployments, strategic and tactical surges, command, control, and communications (C³), ground and airborne missions, and around-the-clock surveillance operations all require humans to make complex decisions under situational stress during extended hours and atypical times of the day.

Transition of Knowledge

A key factor in supporting operational commands is a need for an effective and efficient means to evaluate and transfer technical knowledge and exploratory developments by the laboratory researcher to the operational commander. For example, in the process of identifying and developing decision-making aids to enhance performance, it is essential that the aids be evaluated by operational crewmembers using performance-based criteria during controlled but realistic operational scenarios under extended missions. The Aircrew Evaluation Sustained Operations Performance (AESOP) facility fulfills the requirement for such a transitional laboratory totally dedicated to the development of individual and crew performance measurement methodology to assess the effects of fatigue and assist in the technology transfer of systems to improve human operator decision making under stress.

Triservice-Supported Facility

The AESOP facility has received triservice endorsement and funding to accomplish its function from a development program sponsored by the Office of Military Performance Assessment Technology (OMPAT) at Walter Reed Army Institute of Research (WRAIR), Washington, D.C. The original impetus for the program was the U.S. Army Medical Research & Development Command's need for a comprehensive, high quality, standardized screening system to evaluate the impact of chemical defense protective drugs on military performance. However, program managers quickly recognized that the up-front heavy investments of facilities, hardware, software, performance measurement methodologies, and standardized testing procedures had wider application. The utility of the AESOP was first established by a study requested jointly by the Army and Air Force Surgeon Generals to evaluate the effects of antihistamines on complex decision-making tasks under sustained operations. The tasking area chosen to evaluate the sensitivity,

reliability, and validity of team performance measures was a C³ mission flown by a team of Airborne Warning and Control System (AWACS) Weapons Directors (WDs). The AWACS team members have a C³ mission similar to ground C³ operations, e.g., Control Reporting Post. Several generic workstations were designed and constructed to incorporate flexibility in reconfiguration and for real-time acquisition of individual and team performance measures generalizable to other missions. A network of interconnected computers generated high-fidelity air defense mission scenarios that placed real-world task demands on the operator. The fidelity of the performance tasks permitted the use of real-world operators and provided immediately credible and valid results to the operational world.

A Unique Command and Control Research Facility

The AESOP facility at Brooks Air Force Base, Texas, is a research tool for psychologists, physiologists, and human factors engineers. Unlike tools which have a primary application and must be forcibly manipulated to accomplish other tasks, this laboratory is uniquely multipurpose and flexible. The laboratory consists of

- four reconfigurable programmable crewstations
- an advanced audio communication network
- computer system networks
- high- and low-resolution consoles
- multiple input devices
- a variety of data collection systems
- scenario development tools
- robust software systems for simulation, data reduction and analysis
- support networks for word processing and data handling
- communication servers to remote sites
- a staff of professionals dedicated to programming, systems analysis, technical documentation, simulation development, experimental protocols, and administrative support.

A 3,600 ft² modular facility houses these capabilities and provides a controlled environment for presenting realistic Command and Control (C²) simulations under conditions of sustained operations.

AESOP as a Complex-Decision Research Environment

This triservice facility is the result of years of planning and development by representatives from the Army, Navy, and Air Force. These representatives recognized the need for a controlled environment to analyze the effects of various stressors on military personnel performing complex team decision-making tasks. The Joint Working Group on Drug Dependent Degradation of Military Performance (JWGD³ MILPERF), now OMPAT at WRAIR, gave early programmatic and technical organization to the project.

Systems Research Laboratories, Inc., developed the facility, hardware and software systems as the prime contractor on Navy and Air Force contracts.

Three design principles guided system development:

- creation of a realistic operational setting,
- development of a multipurpose and flexible experimental tool. and
- capability to control the environment and the subjects.

The crewstations and scenarios simulate the air defense mission of the AWACS C³ platform. *A realistic operational setting is achieved through high fidelity crewstation function, authentic graphics presentations, believable scenarios, experienced personnel playing roles in the simulation, and attention to detail.*

The second design principle was the development of a multipurpose and flexible experimental tool. Hardware, hardware systems, software systems, and the facility are integrations of carefully designed modules. The tasks of replacing, restructuring, and building modules can be combined to reconfigure or expand systems, to integrate novel concepts or devices, or to respond to new research requirements.

There is a wealth of task presentation, data collection, and data analysis methods available to investigators. Voice transmission, subject movement, graphics screens, switch activations, keyboard entries, trackball positions, and a host of other events are recorded and time-stamped for comparative analysis. The capability also exists for electrophysiological data collection, voice recognition, voice synthesis, and speech stress analysis.

The AWACS WDs control friendly aircraft against intelligent enemy aircraft that can recognize, engage, and destroy friendly forces based on realistic algorithms. Simulation pilots access decision support tools on their workstations to aid in flying friendly aircraft and in engaging the enemy.

In support of these systems, human factors engineers of NTI, Inc., have developed, tested and refined a large number of individual and team performance measures. These measures include standard performance tests and unique measures developed within the bounds of C² scenarios.

The third design principle concerned the *capability to control the environment and the subjects*. Research protocols include sleep and activity schedules, light and temperature settings, and food and drug ingestion. The AESOP facility is constructed to support these controls in sustained operations. Subjects can be restricted to designated areas of the facility. Minor alterations in the facility and surrounding rooms will allow housing personnel for extended durations.

AESOP - A Multipurpose Research Tool

The AESOP physical plant (Fig. 1) consists of a 2-story 3,600 ft² facility with the following distribution of space:

1. Lower Floor
 - a. 560 ft² computer room,
 - b. 1000 ft² control/recording/testing areas, and
 - c. 80 ft² storage area.
2. Upper floor
 - a. 320 ft² conference/briefing/planning area,
 - b. 360 ft² administrative space convertible to laboratory space,
 - c. 200 ft² data analysis area,
 - d. 850 ft² administrative area, and
 - e. 230 ft² system operations.

The facility has self-contained environmental controls and fire protection, high quality, low noise, controllable lighting, and modular walls.

The four C³ crewstations are high fidelity systems configured as AWACS WD consoles. These consoles have high resolution graphics displays, modular switch panels with programmable switch function, communication panels, QWERTY keyboards, and trackballs. Several high fidelity, low resolution video terminals serve as consoles for simulation pilots, ground controllers, and investigators.

The AESOP computer systems (Fig. 2) consist of:

1. A cluster of 2 VAX 11/780s, 2 MicroVAX IIIs, and a VAXstation III/GPX;
2. Four high resolution, color graphics Silicon Graphics 4D/50 workstations; and
3. Multiple disk drives, tape drives, and printers.

Data recording devices (Fig. 3) include the following:

1. Analog-to-digital and digital-to-analog devices;
2. Sixteen channel FM recorder using VHS format;
3. Low light level video cameras;
4. 1,280 x 1,024 to NTSC resolution scan converter; and
5. Video recorders.

A 10-node audio communication network provides audio communication during simulations. This network is comprised of:

1. Two nodes of computer-controlled voice synthesis;
2. Eight nodes consisting of 8 communication channels each;

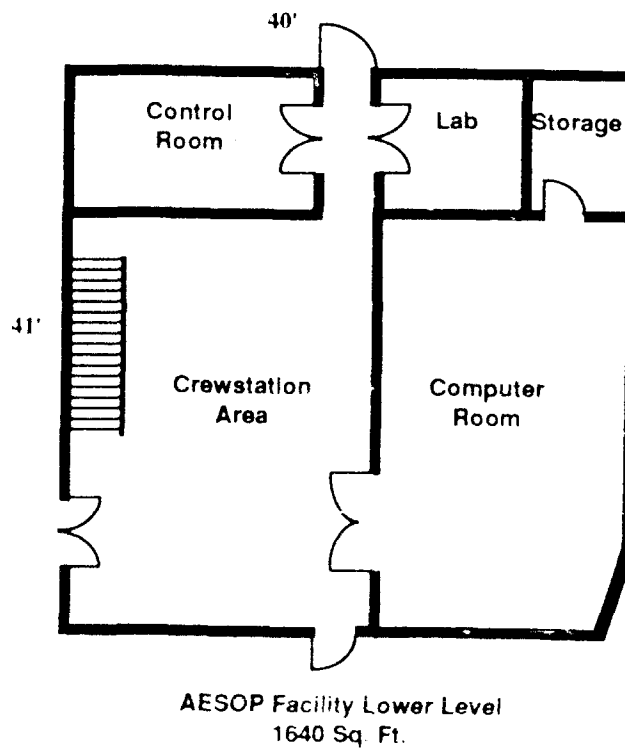
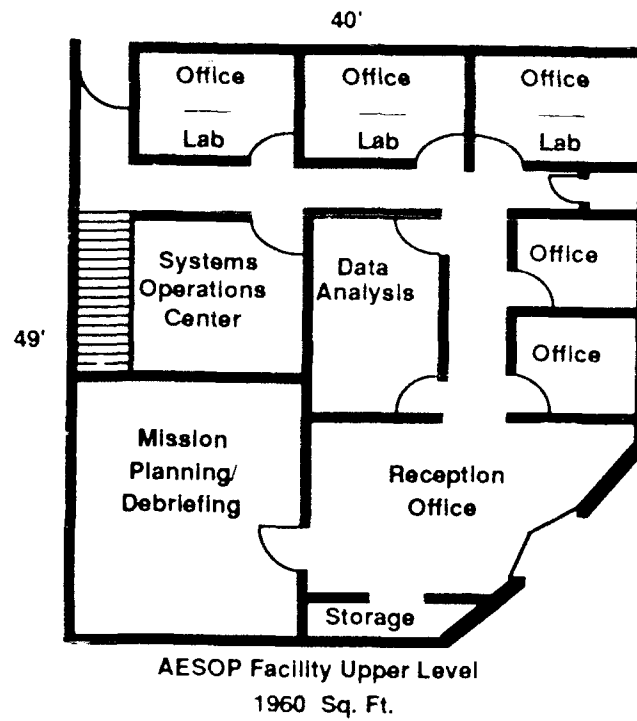


Figure 1. AESOP facility floor plan.

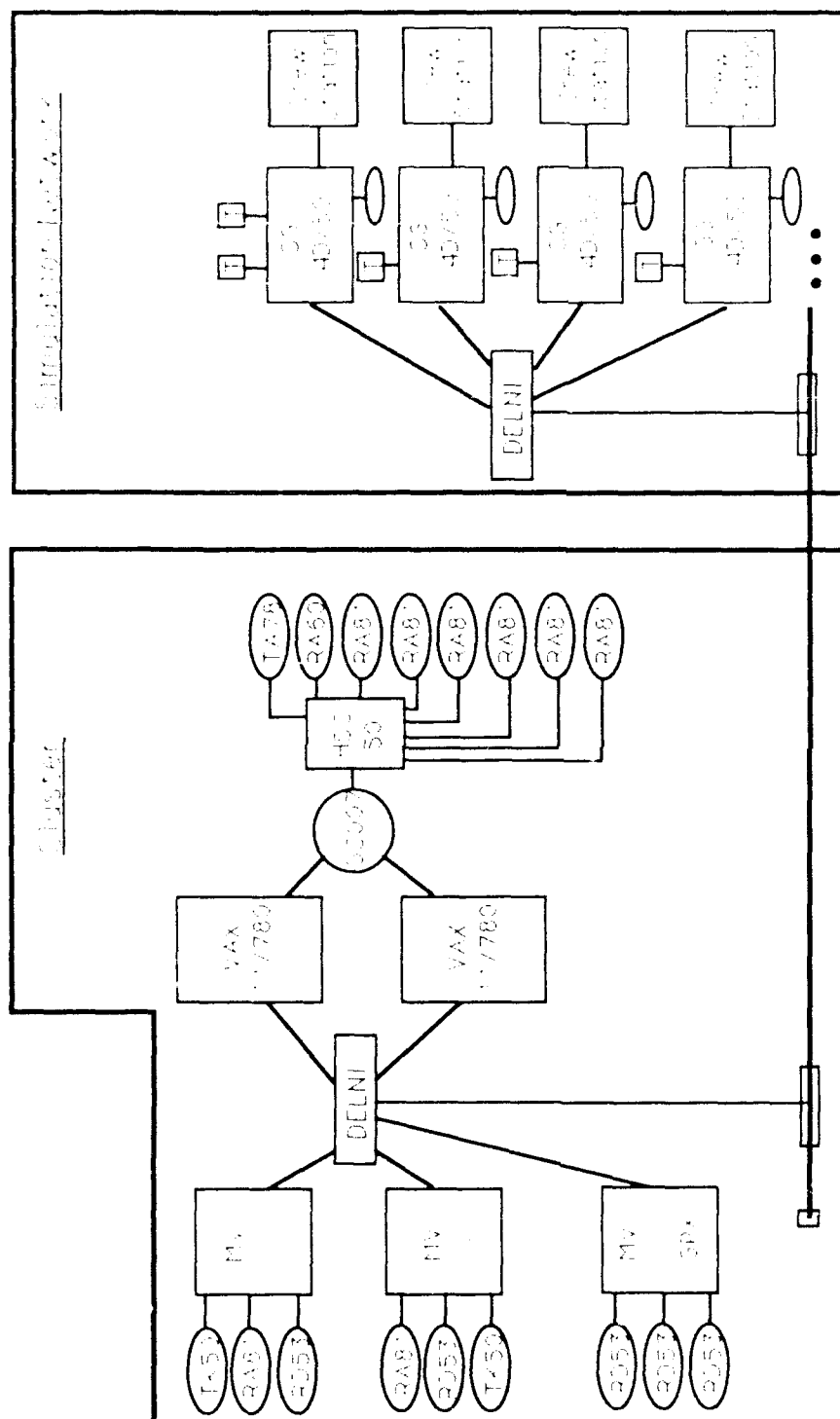


Figure 2. Cluster and Simulation Network.

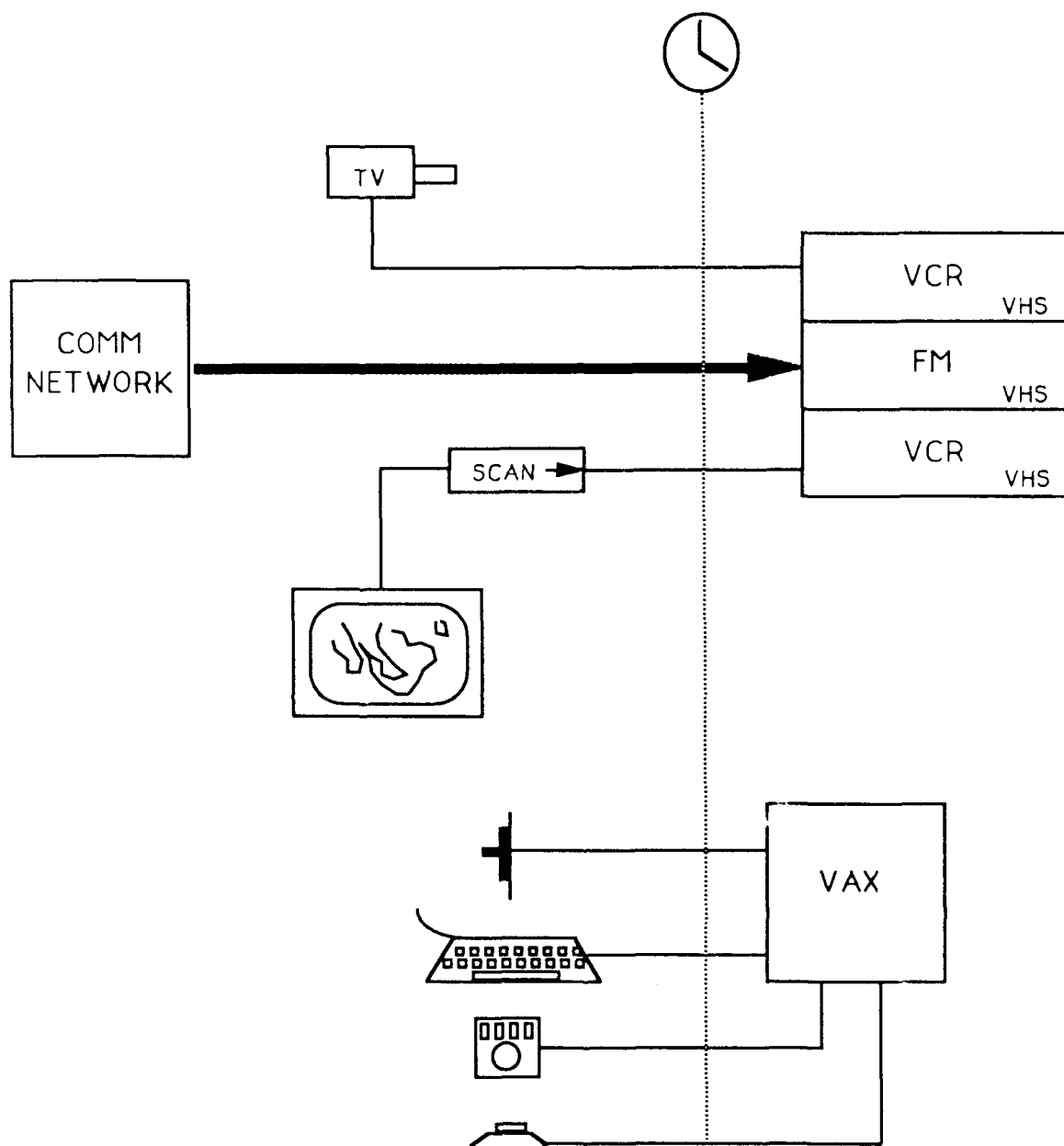


Figure 3. Data recording devices.

3. A central control hub for assigning channel access and connection; and
4. A recording interface for up to 13 channels of audio.

Four additional support systems are available to the staff:

1. Local and long-haul network nodes (Fig. 4) for intra- and interlaboratory access;
2. An integrated microcomputer network;
3. Statistical, database, spreadsheet, word processing, presentation display, and electronic mail systems;
4. Scenarios:
 - a. Seven 3.5-hour air defense scenarios;
 - b. Scoring algorithms for performance measures;
 - c. A range of workloads and stress events;
 - d. Scenario generation, testing, and evaluation capability;
 - e. Voice scripts; and
 - f. Briefing materials.

Applications

The recent completion of a research protocol established the utility and reliability of the systems within AESOP. Twelve crews composed of 3 WDs each from Tinker Air Force Base, Oklahoma, were evaluated in the facility. Each crew tested for 4 consecutive 16-hour days after ingesting either the antihistamines Benadryl or Seldane, or a placebo control. The recently completed tests validated the AESOP facility as a truly flexible research tool for measuring individual and team performance in a sustained operations environment.

Performance-Based Measures of Individual and Team Complex Decision Making

This interim progress report evaluates the sensitivity of selected C³ and synthetic performance measures to the effects of two antihistamine medications, Benadryl and Seldane. Only C³ performance measures are discussed here. A second objective of the research was to assess the magnitude of individual and team performance impairment produced by the antihistamines during high- and low-workload C³ scenarios.

Although the primary goals of the study were drug related, the researchers used the opportunity to gather data on several other issues from several perspectives. These data included: the evaluation of sustained operations and fatigue, the assessment of cognitive workload through embedded tasks, the assessment of stress, the assessment of learning effects, the evaluation of tests for WD selection, and the prediction of complex task performance from cognitive skills status.

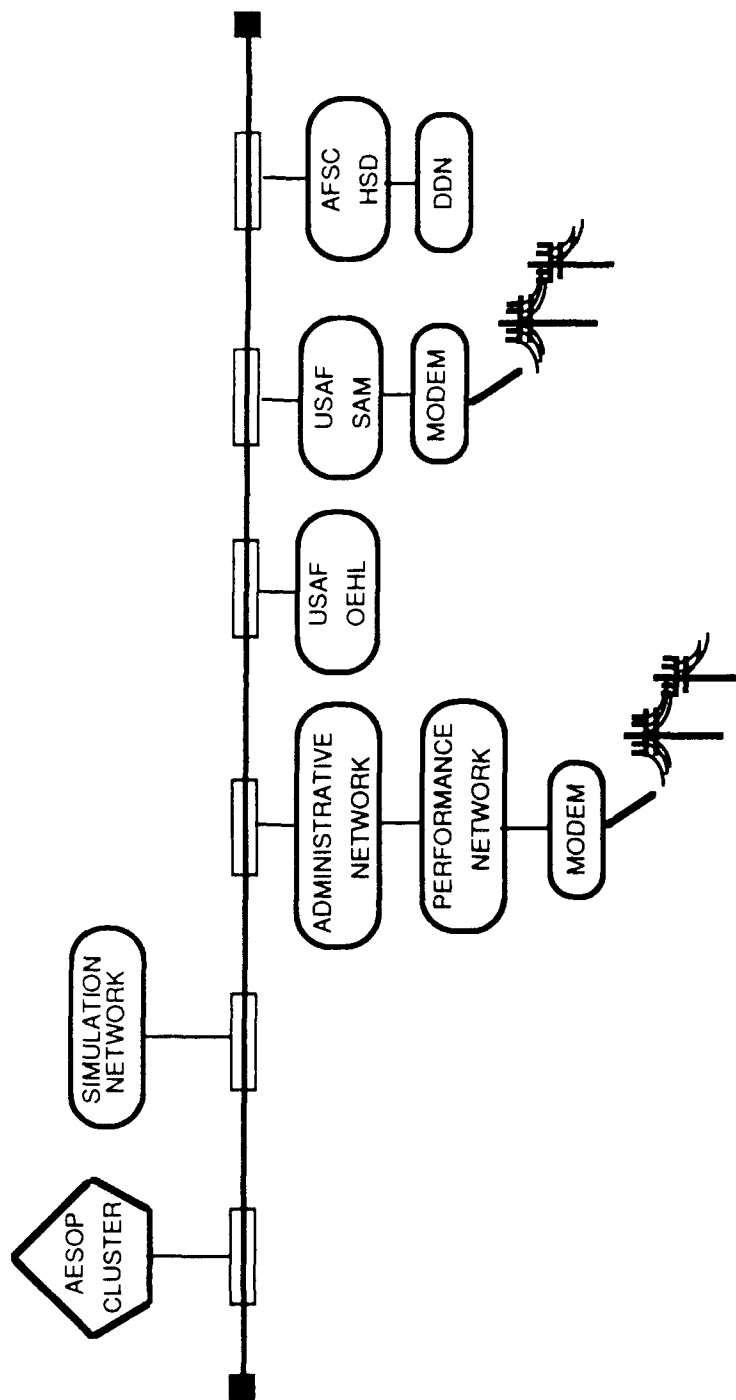


Figure 4. AESOP associate networks.

The Task

WDs in an air defense scenario use their consoles to accomplish a number of tasks. The wartime tasks include locating and identifying aircraft, maintaining track information on aircraft and targets, updating target information received from pilots, accepting aircraft handoffs, performing a tactical controller function with appropriate level of control, providing target briefings to interceptors, performing a tanker controller function, providing recovery assistance, safe passage monitoring, briefing the senior director of any tracking or sensor data problems, and responding to alerts, alarms, and messages on the console. The success of the C³ mission results directly from the WDs' successful accomplishment of their duties.

The WD's goal in a defensive counter air mission (DCA) is to defend friendly lines of communication, protect friendly bases, and support friendly land and naval forces while preventing the enemy from carrying out offensive operations. The primary operations are conducted to detect, identify, intercept, and destroy enemy aircraft attempting to attack friendly forces or penetrate friendly airspace.

Performance Measurement

Obviously the performance of such a complex system including human operators depends on numerous interacting internal and external factors. Because of these multiple determinants and numerous data perspectives, it was necessary to use a variety of metrics to characterize the system and to diagnose the sources of observed variations in system performance. Few people would argue that any one measure is sufficient by itself to characterize a complex system's performance; most would agree that additional measures add something to the measurement process. The interpretation of large metric sets is facilitated by an implicit underlying structure that weights the significance of each measure and relates it to the others.

Hierarchy of Performance Measures

One way to approach understanding the relationships among measures is to devise a hierarchy of performance determinants that will provide a classification framework for individual measures. Each level of such a hierarchy would contain groups of measures that jointly determine the measures available at the next level higher in the framework. This was our approach in studying AWACS WDs conducting an air defense scenario in our laboratory.

We chose four levels of measures:

- Mission Effectiveness,
- System/Team Performance,
- Individual Performance, and
- Performance Capability.

Figure 5 illustrates this hierarchy graphically; Figure 6 expands upon the diagram.

The highest level of the hierarchy contains indices of *Mission Effectiveness*. These measures are derived directly from the specific objectives of the mission assigned to the system. An example would be protection of a specific sector of air and ground space from infiltration by enemy aircraft (protection of assets). Measures that flow from such a high level objective and that would assess performance in terms of mission effectiveness might include the following: number of enemy infiltrations, amount of fuel and weapons expended, and ratio of enemy lost to friendly assets.

The second level of the hierarchy, *System/Team Performance*, contains those groups of measures that reflect factors immediately affecting mission effectiveness. Such measures of System/Team Performance reflect the degree to which the combined man-machine system has accomplished those tasks required to meet mission objectives. These metrics do not reflect the individual contributions of different human behaviors. Instead, they are more global indices of the degree to which the total system successfully accomplished the tasks essential to mission success. For example, the WD/workstation system represented in the C³GW is required to meet its mission objectives essentially by accomplishing an air traffic control task aimed at directing interceptor aircraft to defeat threat aircraft. This air traffic control task may be broken down into a number of essential subtasks such as pairing interceptors with targets and providing target data to interceptors. The *System Performance* tasks would be the same or similar to the tasks that would be performed if the AWACS was on a different mission, e.g., *offensive* counter air. Performance measures of the latter include the accuracy and speed of data transfer to interceptor pilots.

The third level of the hierarchy is comprised of specific groups of measures that assess the individual contributions of human components to overall system performance. Measures included in the *Individual Performance* level of the hierarchy reflect the quality of the individual behaviors required of the WD expressed primarily in terms of latencies and errors. These metrics are derived by examining the system functions required to meet mission objectives to identify the specific contributions of the operator. For example, the system performance requirement to pair targets with interceptors may require the WD to identify a target's location on the workstation display and communicate this information to an interceptor aircraft via radio. The quality of the operator's performance in achieving this objective might be measured by evaluating the time needed to complete the full sequence of required behaviors and by assessing the accuracy of each manual and verbal response. This sequence can be broken down into the different components of the DCA mission: detection, identification, interception, and

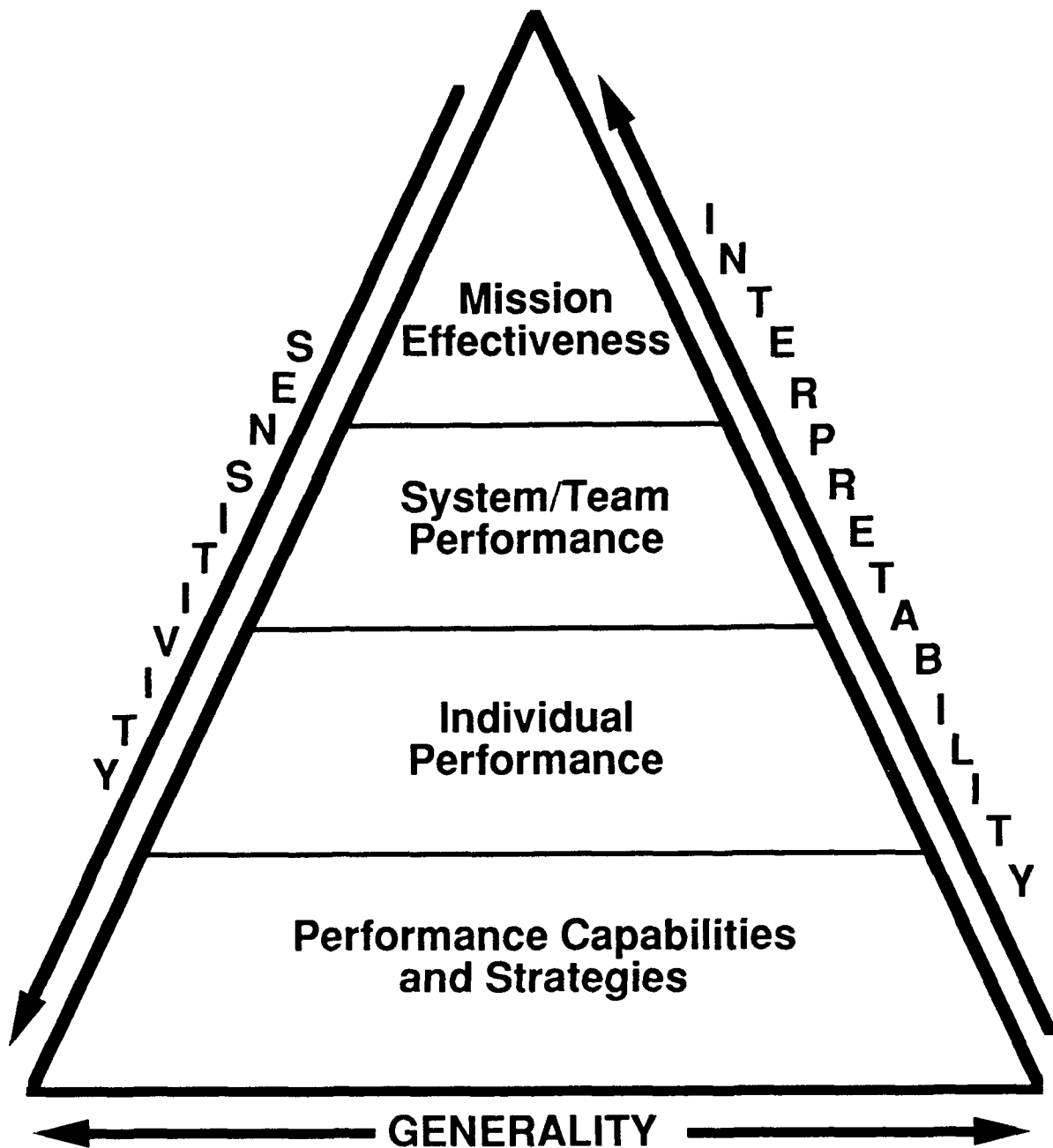


Figure 5. Performance measurement hierarchy.

MISSION EFFECTIVENESS: Air Defense

Objective 1: Protection of Assets

Measure 1: Number of friendly airbases damaged

Measure 2: Number of hostile aircraft destroyed

SYSTEM / TEAM PERFORMANCE: Weapons Control

Objective 1: Pair interceptors with targets

Measure 1: Ratio of pairings to targets

Measure 2: Ratio of provision of target data to requests

INDIVIDUAL PERFORMANCE: Communications

Objective 1: Communicate target information to interceptor

Measure 1: Time to complete communication

Measure 2: Accuracy of manual and verbal responses

PERFORMANCE CAPABILITIES: Workload

Objective 1: Preserve processing capacity

Measure 1: Response time on embedded task

Measure 2: Number of omitted low priority tasks

Figure 6. Levels of performance measurement.

destruction. Over 130 measures of this type were tabulated for data reduction and analysis.

The final level of the proposed hierarchy contains metrics that assess factors directly affecting the individual performance capacities of primary system components. For the human operator, measures of *Performance Capability* are composed of a large group of potential human state and ability metrics that combine to determine overt performance. These metrics include indices of workload or reserve processing capacity; fatigue; arousal level; experience level; personality; and individual perceptual, cognitive, and motor abilities.

The following tests and questionnaires were included at this level: a biographical sketch, a WD experience form, personality scales for potential use in developing WD selection tools, and surveys of their current state (symptoms, sleepiness, fatigue, etc.). The scales included the Rotter Scale, which assesses the locus of control generally perceived by a person in causing changes to take place in one's life; the Personal Characteristics Inventory (PCI), which assesses attitudes and leadership qualities; the Life Style Questionnaire, which predicts a subject's performance under stress, the Least Preferred Co-worker (LPC) Scale, which may identify a WD's leadership style; the Jenkins Activity Scale, which assesses a WD's personality characteristics of decision making; the FIRO-B, which measures a subject's attitudes with regard to sociability and social interaction; the USAFSAM Fatigue Scale, which allows the subject to describe feelings at a particular time; an Operational Impact Survey, which allows a subject to rate how well the team completed its mission and how well each subject contributed to the mission; a Scenario Evaluation form allowing each WD to order the simulations with respect to workload; and the Subject Workload Assessment Technique (SWAT), which allowed each subject, at the end of each simulation, to evaluate the difficulty of the scenario along SWAT's three dimensions: time load, mental effort, and psychological stress. The WDs kept logs similar to those kept during a standard mission. They recorded aircraft call signs, type aircraft, target numbers paired against, check-in time, weapons states on the aircraft at Return to Base (RTB), results, and other information.

Performance Capabilities extend along a dimension from states and traits highly resistant to change, such as personality characteristics, to those that change from moment to moment, such as reserve processing capacity. Some can, therefore, be considered potential performance predictors while others are dependent measures.

The multi-level classification of performance measures previously proposed has the advantage of placing metrics into logical subordinate and superordinate groups indicating the predictive relationships among them. Measures at each of the levels differ in their sensitivity, generalizability, and practical interpretability. *Mission Performance* level measures are easily interpreted, but not generalizable to other types of scenarios. *Cognitive Capabilities* are highly generalizable to different types of tasks and scenarios, but difficult to relate directly to mission objectives.

Problems In Measuring Performance In a Complex, Two-sided Environment

Kubula (1978) described many of the problems in attempting to measure performance in a 2-sided test. Although the realism of an aggressor force adds to the reality of the scenario, it also makes each test unique. Some of the problems include: the non-repeatability of events from one team to the next, allowing one team to overextend themselves on one problem such that they are not ready for the succeeding events programmed into the script, and responses that are unique to only one team and hence cannot be compared to the responses of other teams.

We solved some of the problems in our simulations by having a single Senior Director (SD) who was a part of the experimenter's team of players. The SD kept the team in bounds by conserving enough resources to fight the war and by breaking off intercepts and other distractions that would remove the WD from important upcoming events requiring specific responses. These "assists" by the SD will be counted against the team in some weighted manner.

Methods

The 552d AWACS Wing assigned 12 teams of 3 WDs (male and female), who previously volunteered, to spend their work week in support of this study (Fig. 7). The teams were randomly assigned to a drug treatment condition and either low-high or high-low workload order of scenarios.

The WDs arrived on Saturday or Sunday evening for a preliminary briefing. Training took place on Monday for about 8 hours. Teams received training on 6 simple computerized tests and 2 complex tests over approximately 4 hours. They also ran a 3-hour C³ training scenario to familiarize them with the simulated AWACS crewstations and scenarios. Due to the difference in the appearance of the drugs, subjects ingested 1 Benadryl placebo and 1 Seldane placebo at 2230 or before going to sleep.

Starting on Tuesday, the teams were tested in two 3.5-hour scenarios each day for 3 days (Fig. 8). Each group ingested only placebos during the testing schedule for Tuesday. A randomly assigned team ingested the recommended therapeutic dose of either Benadryl, Seldane, or a lactose placebo starting on Tuesday evening. Total antihistamine/placebo ingestion for each group consisted of either eight 25-mg Benadryl, four 60-mg Seldane, or all placebo preparations.

Additional Measures

Besides the outcome measures of how well a team or individual performs in a simulated air defense scenario, it is important to understand the underlying processes that contribute to those outcomes. Embedded tasks were used to measure *reserve capacity*, *team coordination*, and *situational awareness (SA)*. These tasks are natural to

Subjects:	12 teams of 3 Weapons Directors from 552d AWACS Wing, Tinker AFB. One team each week.
Drugs:	Benadryl (25mg), 4 per day Seldane (60mg), 2 per day Placebo
Training Day:	1
Testing Days:	3
Scenario Difficulty: (each Day)	1 Hard and 1 Easy

Figure 7. C³ antihistamine study.

DRUG	DAY 2	DAY 3	DAY 4
Benadryl	Easy Hard	Easy Hard	Easy Hard
Seldane	Easy Hard	Easy Hard	Easy Hard
Placebo	Easy Hard	Easy Hard	Easy Hard

Figure 8. C³ research design.

the air defense scenario, but are low priority tasks. These tasks were delivered by voice queries articulated by the Votan speech synthesizer or vocally by the SD. The 3 embedded measures for reserve capacity are:

1. Whether or not a response is given.
2. Accuracy of the responses.
3. Latency of the responses.

The independent variables that may determine the WD's workload level are: the number of flights currently under the WD's control, the level of control of each flight, the Air Defense Warning Level (ADWL), and the number and type of additional tasks currently being worked by the WD. A typical SD query for reserve capacity might be "What state armament/fuel on the aircraft under your control?" Low workloads should result in quick, accurate responses from the WD. High workloads should result in ignored requests, partial information, and long response times.

Individual members of a WD team can work independently of each other. However, since the enemy is directing the attack in an air defense scenario, the battle does not always unfold the way it is planned in a mission prebriefing. As a result, each WD's responsibilities change throughout the mission. These changes should be adaptive and result from insight and leadership. Further, the adaptations require cooperation and coordination among the team members. The WD responses involve passing and confirming information to each other and accepting responsibility for incoming requests when time is available. Embedded measures for team coordination include:

1. Whether or not the information is passed to the other WDs.
2. Accuracy of the response.
3. Latency of the response.

An event designed to elicit a team coordination response might be an ADWL announcement from ground control.

To deal effectively with events in an air defense scenario, a WD must maintain an accurate representation of the battle. This representation (both internal memory and external notes) defines the WD's awareness of the current situation. If a representation is in error, the WD may commit to kill rather than identify an unknown target. Therefore, throughout the scenario the WD's awareness was probed for the correct ADWL, for the status of airbase openings/closings, and for the status of hot/cold surface-to-air missile (SAM) sites. The embedded measures for situational awareness are the same as for workload. An event expected to elicit a response would be for the SD to tell WD1 to kill track 0304. The WD should question this command during peacetime since the SD had no authority to issue the order.

Cutting across all levels of measurement are strategies used by WDs in setting goals, accomplishing tasks, and evaluating outcomes. These strategies will be coded from the transcripts of the WD's verbal communications records. Adaptive and

nonadaptive strategies will be identified and assessed against performance outcomes at all levels of measurement. Results were not available at the time this report was written.

Development of Realistic and Complex Command, Control, and Communications Scenarios

In some respects, a C³ environment is synthetic. Information about the real world is gathered and translated into symbols. These symbols are presented systematically. The C³ operator, after translating these symbols, applying learned principles, and making a decision, then initiates actions, which in turn are translated into symbols and sent back to agents in the real world. Activity occurs, the real world changes, and the cycle repeats. Such an environment can be effectively replicated in a simulation.

C³ Environment Model

A five-part model (Fig. 9) can be constructed to illustrate the C³ environment:

- symbol presentation,
- symbol input,
- agents who initiate response,
- change in stimulus symbols to reflect the response, and
- the C³ operator.

The key to creating a realistic simulator that can be programmed to produce worthwhile data lies in emulating the C³ environment.

Command, Control, and Communications Generic Workstation (C³GW)

The Command, Control, and Communications Generic Workstation (C³GW) is a multiple task simulation system used to develop measurement methodologies to assess team performance in a C³ environment. In our recent research, the technical objective was to generate mission scenarios that place realistic task demands on AWACS WDs in a simulated C³ environment. The C³GW provided mission scenarios, tasks, controls, and displays replicating the functions of the AWACS WD crewstation using methodologies described in Eddy (1989).

Developing the C³GW

To develop the C³GW, we refined two parts of the C³ environment model. We had real-world C³ operators--AWACS WDs. We also developed the heart of the model, a replica of the AWACS WD crewstation. One of the first development issues encountered

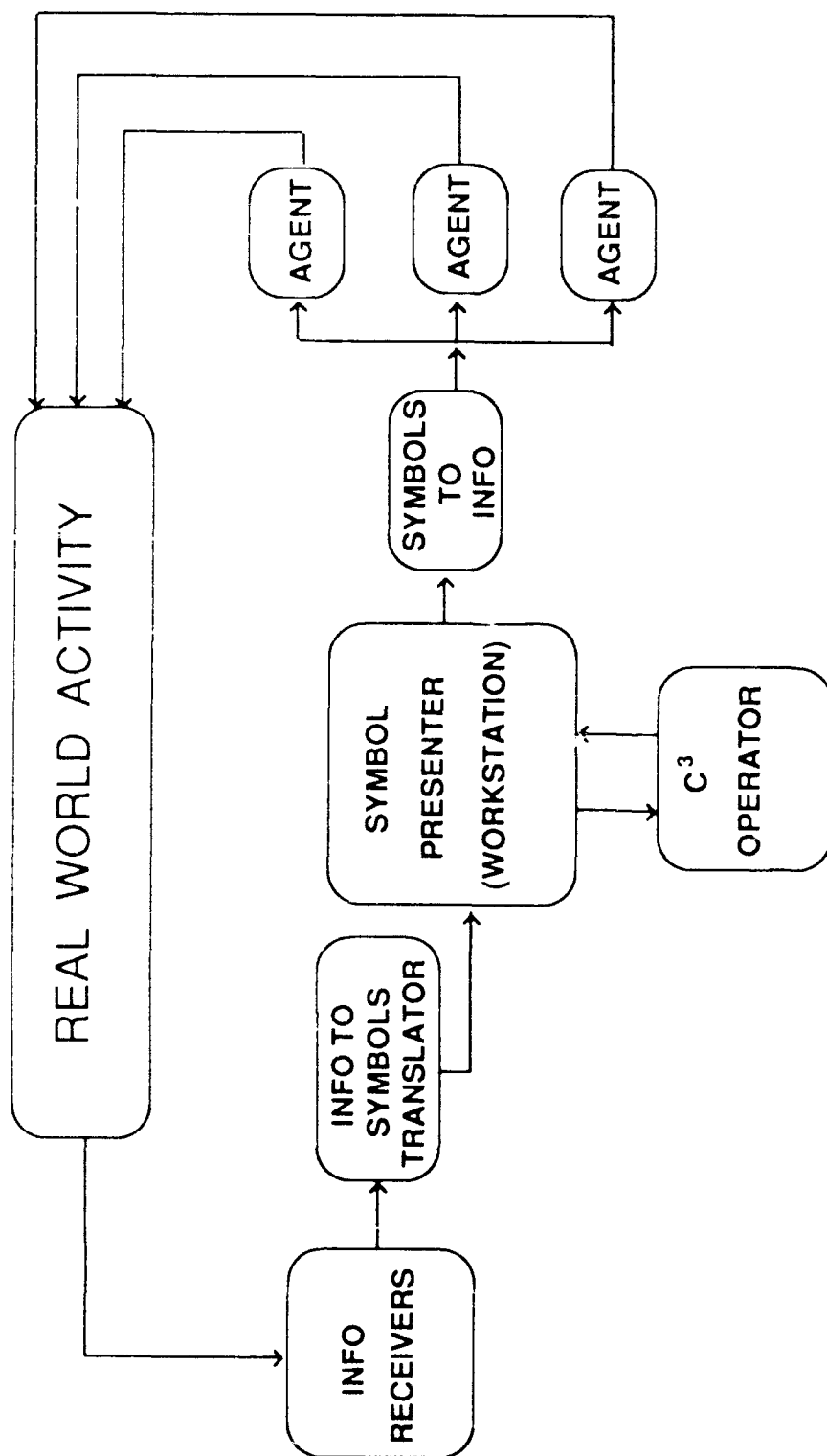


Figure 9. C³ environment model.

was how to make the crewstation authentic without sacrificing the capability to both control the stimuli and to measure AWACS WD responses.

The C³GW is very similar to the AWACS crewstation. Considering the high motivation and intelligence of our subjects we thought that, with a short training session, they would rapidly adjust to the slight differences. Two examples will suffice:

1. We used the standard QWERTY alpha-numeric keyboard instead of the AWACS crewstation keyboard.
2. We developed software for only the most commonly used functions, and then for only those options most frequently used to carry out assigned tasks. On the E-3, when a WD presses the COMMIT switch for example, a menu assist line displays all the possible options. When the COMMIT switch is pressed on the C³GW, the display is the same, but a few of the options are not implemented.

By paying close attention to tradeoffs, developing reasonably close workarounds, and keeping objectives in mind, realism was attained even though the environment was not exactly replicated. The verbal feedback from those AWACS WDs who tested and evaluated the C³GW before our first testing runs, as well as the test subjects themselves, indicated these were valid assumptions.

Scenario Development

In developing and refining scenarios for testing, a DCA mission was chosen because it encompassed the widest variety of WD performance tasks. Using one basic scenario as a template, 7 scenarios were developed. Each scenario fulfilled 6 interlocking and overlapping objectives:

- pushed through different command and control levels,
- created task loading,
- delineated some of the "embedded tasks" for measuring responses,
- appeared to be different from the others while structured the same,
- established levels of difficulty, and
- used unclassified information.

These objectives relate to the part of the model dealing with the symbols that act as stimuli.

Different command and control levels require WDs to respond and perform in different ways. For example, to intercept an unknown track during peacetime, the WD uses stern intercept geometry for visual identification. At a higher command and control level, cutoff geometry is used. During wartime, the WD uses the best geometry based upon position, fuel, armament, and tactics for target destruction. All 7 scenarios use the same 6 command and control levels, or Air Defense Warning Levels (ADWLs). These

scenarios are modeled after NORAD's DEFCONs. In this instance, we gave up a little realism for a larger variety of responses to measure. It is rare indeed to go from peace to war in under 3 hours.

Task Loading

To create a training scenario, 3 easy scenarios, and 3 difficult scenarios, we varied the task loading (Fig. 10). The scenario task loading falls into 4 distinct categories:

- nuisance-scripted events,
- noncritical events,
- resource diverting events, and
- mission critical event information.

Nuisance-scripted events, such as civil air traffic control chatter on guard, distract the WD's attention.

Noncritical events don't directly affect the WD's primary accomplishment of the DCA mission but they do make mission conduct smoother. Some examples of noncritical inputs are low-level altimeter readings, checking of Identification Friend or Foe/Selective Identification Feature (IFF/SIF) squawks, and the passing of radar height variance information. Both nuisance-scripted inputs and noncritical inputs help generate the "fog of war" adding to the realism of a scenario.

Resource diverting events force the WDs to use their interceptor resources for tasks unrelated to the primary DCA mission. Examples of resource diverters are responding to a civil aircraft emergency or escorting a defector or a medical evacuation (medevac) aircraft.

Mission critical event information changes the environment and the WD's mode of execution. Examples of mission critical information are air base openings and closings, changing the area of responsibility, and friendly SAM site activations.

The foregoing types of task loading allow a subtle form of experimental control over the stimuli and types of responses to capture. Task loading proved significant, for example, in affecting WD output performance in preventing total numbers of hostile aircraft penetrating to target.

Embedded Tasks

Embedded tasks are similar to nuisance-scripted inputs. Some examples include SD requests for a WD-controlled aircraft's Mode 4, fuel state, or armament state. They indirectly measure a WD's capacity for work. Embedded tasks are the kinds of tasks a WD must perform to conduct the DCA mission.

Nuisance-Scripted	Noncritical	Resource Diverts	Mission Critical
ATC Chatter	Low Level Altimeters	Hijack	Hostile Tgts
SD Enquiries	Checking IFF/SIF	Defector	Airbase Open/Close
--Mode 4 Results	Passing Radar Ht	Escorts	AOR Chg
--Fuel States	Var Info	--Medevacs --VIPs	SAM Site Active
--Armament States		A/C Emergencies	ADWLS

Figure 10. Weapons Director task loading events.

Standardization of Design

All 7 scenarios have the same basic design. This design allows for cross-checking among scenarios. Each scenario is based on a standard enemy attack of 4 waves. The first wave is a reconnaissance probe and has only 3 enemy aircraft. The next 2 waves have 12 enemy aircraft each. Most of the attackers are bombers escorted by fighter/bombers. The last wave has a mass of 16 enemy aircraft; usually one bomber escorted by 2 fighter/bombers.

The course of each attacker is laid out on an XY-coordinate plane. A latitude/longitude map is overlaid on the XY-map so each (x,y) corresponds to a lat/long point. To make each scenario appear unique, the x,y plane is rotated a number of degrees, and matched with a lat/long center from a different geographical region.

To get a good fit of the geographic points for the enemy and friendly bases, some of the XY coordinates were changed. This change added to the realism, but detracted from the control since the background basic inputs were not the same. The following chart gives an indication of how the basic inputs were developed.

NAME	TYPE	BASE	ROTATION	SCALE
Yakima	Training	Old Troy	165° Right	1:21800000
Troy	Low Workload	Old Troy	28.5° Right	1:5400000
Canaan	High Workload	Troy	118.5° Left	1:2700000
Thebes	Low Workload	Troy	151.5° Right	1:5400000
Uruk	High Workload	Canaan	135° Right	1:5400000
Gomeria	Low Workload	Troy	125° Right	1:5400000
Aaragon	High Workload	Canaan	180° Right	1:5400000

Scenario Differentiation

To further disguise the scenarios from each other, names and call signs were changed. The timing of some of the task loading was also changed. For example, a medevac flight requesting escort appears 15 minutes into a scenario, while in a similarly task-loaded scenario a VIP flight requesting escort appears 17 minutes into the scenario. The idea was to make each of the essentially equivalent scenarios seem fresh and different to the subject WDs. Yet another factor is the use of slack time between each major stimulus. This factor allowed experimenter intervention to set up the same prior condition for each new stimulus, if necessary.

Scenario Refinement

The most difficult part of the scenario development is the orchestration of visual input stimulus and aural input stimulus. To be coherent and believable, these stimuli

have to mesh together. For instance, if you want a hijacked aircraft to fly through the operational area, the radar dot symbol with its accompanying IFF/SIF symbol has to start somewhere and fly a route. Both symbols must be credible given the context of the scenario. To make sense the visual dots must precede the aural guard transmission. In addition, the other stimuli events must be taken into account. Questions as to whether the stimulus must occur on its own or simultaneously with other inputs must be answered.

Running the Simulation

To address that part of the model referred to as agents, we brought in retired USAF pilots. Since some of the responses required of the simulation pilots could not be perfectly predicted, the retired Air Force fighter pilots acted as interceptor simulation dot drivers. They responded to WD directives and queries in a real-time fashion and in a realistic manner.

Due to space and equipment limitations, each pilot simulated more than one pilot/airplane at a time. To prevent task-saturation, to add realism, and to retain control of the experiment, the friendly fighters were given some automated parameters. If, for example, at 21 nm the parameters were met, the computer system took over control of the fighters and flew in the final attack phase for the simulation pilots. A JUDY alert was given to the simulation pilot to transmit, as well as weapons launch of FOX 1, for example.

There were also lethal engagement parameters. Tracking of fuel and armament consumption was automated with the appropriate BINGO and WINCHESTER calls. Since this simulation was unclassified, the parameters were developed from open sources with an eye on following principles, such as burning more gas at faster speeds and at different altitudes. A similar set of parameters was set for hostile fighters. This set ensured that the threat automatically emulated a real-world threat of engaging and destroying friendly aircraft. Although constraints did not allow for a highly sophisticated emulation, the advantage of control was achieved because the parameters must always be met for engagement, or it doesn't happen. Yet the enemy does appear to respond. So the model is brought full circle back to the status of the symbols being presented.

State-of-the-Art Research

Simultaneously, the Aerospace Research Branch provided access to a state-of-the-art research facility designed with the flexibility to introduce advanced theoretical constructs into C³ scenarios with embedded performance measures. The existing database on AWACS team performance was structured to encourage scientific inquiry and to answer relevant operational questions. For example, the Human Resources Laboratory's C² Ground Operations Branch at Wright-Patterson Air Force Base (WPAFB), Ohio and the USAF School of Aerospace Medicine's Sustained Performance Function jointly verified an operator model of tactical decision making to evaluate the impact of automation on information processing. Near-term cooperative studies will further exploit

commercially available relational database management tools to ask "what if" questions for applications in crew training, composition, cohesion, and autonomy in a distributed decision-making network. Mid-term studies will evaluate naturally occurring sleep promoters, performance enhancing stimulants, and bright-light treatment for resynchronizing disrupted circadian rhythms in operational crews.

The mid-term development efforts will conclude with a series of studies that require real-time monitoring of electrophysiological signals to establish the arousal state of the operator under low tasking situations. The results from these research studies will provide a database for model development of operator fatigue to advise operational commands on work-rest shift cycles.

Future Performance Research

The balance of operational relevance and flexibility of the AESOP facility will make it a productive transitional laboratory for several years. To extend the longevity of the AESOP facility, future projects will connect to other Department of Defense (DOD) C³ simulation facilities via long-haul networks to provide the capability to run multi-mission scenarios, e.g., Simulation Network (SIMNET). Future areas of research will investigate algorithms for improving the validity of human operator models by degrading performance over time. The enhanced models will then be applied to the design of crewstation interfaces using neural networks and artificial intelligence methodology to maximize task allocation between operator and machine.

Recommendation

The Sustained Operations Branch considers the initial heavy triservice investment in AESOP as a long-term strategy to provide a gateway for systematically transferring advanced technology to the operational commands. For the DOD to achieve full benefit from this resource, we should coordinate work in areas of general interest that go beyond a simple information exchange to avoid duplication of research efforts. If the DOD research community recognizes a common need for this type of technology conduit for transferring performance-based measurement methodologies, then the AESOP facility and existing database should be made available for broader applications on a cost-sharing basis. The sharing of databases among participants in cooperative programs has the effect of a force multiplier yielding a greater return on the resource investment by an individual organization. In this age of austere budgets and a diminishing government work force, this is only a common sense approach. Cooperative agreements exist with the National Aeronautics and Space Administration (NASA) and are actively being pursued with the Federal Aviation Administration (FAA) and other DOD divisions.

References

- Eddy, D. R. (1989). Selected team performance measures in a C³ Environment--An Annotated Bibliography. Technical Report USAFSAM-TR-87-25, USAF School of Aerospace Medicine, Brooks AFB, Texas.
- Kubula, A. L. (1978). Problems in measuring team effectiveness (Report No. HUMRO-PP-2-78) Alexandria, VA: Human Resources Research Organization (DTIC No. AD A049 560).