Report No. FAA-RD-75-186

DEVELOPMENT OF A PERFORMANCE CRITERION FOR ENROUTE AIR TRAFFIC CONTROL PERSONNEL RESEARCH THROUGH AIR TRAFFIC CONTROL SIMULATION: EXPERIMENT I - PARALLEL FORM DEVELOPMENT

E. P. Buckley, et al.

FEBRUARY 1976

INTERIM REPORT

Prepared for

U. S. DEPARTMENT OF TRANSPORTATION
FEDERAL AVIATION ADMINISTRATION
Systems Research & Development Service
Washington, D. C. 20590

NOTICE

Technical Report Documentation Page

1. Report No. FAA-RD-75-186	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle DEVELOPMENT OF A PERFORMANCE CRITERION FOR ENROUTE AIR TRAFFIC CONTROL PERSONNEL RESEARCH THROUGH AIR TRAFFIC CONTROL SIMULATION: EXPERIMENT I--PARALLEL FORM DEVELOPMENT.		5. Report Date Feb[redacted] 1976
		6. Performing Organization Code
7. Author(s) Edward P. Buckley, et al.*		8. Performing Organization Report No. FAA-NA-75-40
9. Performing Organization Name and Address Federal Aviation Administration National Aviation Facilities Experimental Center Atlantic City, New Jersey 08405		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. 122-112-100
12. Sponsoring Agency Name and Address U.S. Department of Transportation Federal Aviation Administration Systems Research and Development Service Washington, D.C. 20590		13. Type of Report and Period Covered Interim Jan[redacted]—Fe[redacted]75
		14. Sponsoring Agency Code

15. Supplementary Notes

* Bernard Goldberg, Richard Rood, Helen Hamilton, Florence Champion, James Taiotta, R. Algeo, S. R. Tszczolkowski, Gerard Spanier, Francis Baldwin, Jacqueline Ludel, and Edward Stevens.

16. Abstract

The first of a series of small experiments was performed as part of the process of developing a standardized performance criterion for journeyman enroute traffic controllers. The finally developed performance measurement system will be used in personnel research such as the evaluation of potential aptitude tests as to their capacity to predict suitability for entrance into training.

The criterion measure will be based on the use of realistic dynamic simulation of the radar air traffic control situation. The completed measurement system will be required to possess reliability, objectivity, and relevance of measurement of performance. Another requirement will be the availability of alternate traffic problems which are different but proven to be of equivalent difficulty level.

The purpose of this first experiment was to seek directions for the construction of different but equally difficult (parallel) forms of the test by using combinations of sector geographic structures and traffic density levels. Two sectors, which differed widely in geographic structure, and three traffic density levels were orthogonally combined to yield six experimental conditions. Six experienced air traffic controllers worked under each of the six conditions in the air traffic control simulator. The results indicated that performance scores were much less affected by sector structure than by traffic density. Consequently, it was accepted as a guideline for further work that parallel forms can be built on the basis of traffic density level equivalence alone. This will simplify development of parallel forms of the criterion measure.

17. Key Words	18. Distribution Statement
Personnel Research Air Traffic Control Performance Criteria	Document is available to the public through the National Technical Information Service, Springfield, Virginia 22151

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	62	

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

PREFACE

Sincere acknowledgement is made of the important contrubutions of Mr. Bernard Goldberg and Mr. Richard Rood who designed the traffic samples and conducted the runs and made many substantial contributions of several kinds.

Acknowledgement is made of the contribution of Mrs. Florence Champion and Mr. J. Tallota who collected, collated, and reduced the data, and worked on the analysis of the data from inception to completion.

Acknowledgement is made of the contribution of Dr. H. Hamilton who conducted the analysis of the SRI indexes presented herein.

Acknowledgement is gratefully given for the work of the members of the Systems Development Branch at NAFEC who developed and operate the digital air traffic control simulator. Particularly important contributions were made in this particular experiment by Mr. Stanley R. Pszczolkowski, who developed a method of using digital data to simulate operational broadband radar displays, and to Mr. Richard Algeo, who developed the real-time performance measurement scoring system.

Acknowledgements and thanks are rendered to Mr. Gerard Spanier and Mr. Francis Baldwin who designed a heart rate measurement device utilizing existing components which could conveniently be used in the simulator.

The professional skill of Dr. Jacqueline Ludel in exploring background materials and literature is gratefully acknowledged.

The assistance of Mr. Edward H. Stevens of Computer Sciences Corporation who contributed the bivariate symmetry test statistical analysis is gratefully acknowledged.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

INTRODUCTION

BACKGROUND.

The experiment being reported herein is one of a series of small experiments having the overall objective of developing a criterion measurement system appropriate for the position of enroute air traffic control specialist in the Federal Aviation Administration (FAA). The criterion measurement system which is being developed will be hereafter referred to as the CPM, for Controller Performance Measurement system. It will be based on the use of dynamic real-time simulation of the air traffic control system.

Dynamic air traffic control system simulators are usually used for equipment and system evaluations and comparisons. They have only once, it is believed, been used to objectively measure individual controller performance, prior to the experiment being reported upon here. That previous experiment was reported upon in 1969 by National Aviation Facilities Experimental Center (NAFEC) (reference 1).

The uses to which such a measurement system could be applied are many and varied. One of the more urgent needs it could fill is that of an objective performance criterion measure against which to validate (i.e., determine the predictive ability of) aptitude tests for air traffic control personnel. (For a discussion of the history of aptitude testing in air traffic control, as well as the other areas in which criteria are needed, see reference 2.)

In order to be used for any purpose, certain characteristics and options must be demonstrably present in the finally developed system. Among these are content validity, test-retest reliability, and the availability of parallel forms. (For a discussion of these and other requirements to be met in criterion measure development, see reference 3.)

PURPOSE.

The particular experiment being reported upon here had the purpose of exploring one method of constructing parallel forms. Parallel forms of a measurement system are "editions" of the test which cover the same substance but with different material (e.g., items, questions) and are of approximately equal difficulty. The purpose of parallel forms is to make available different, but equal, tests should retesting be required, and also to prevent the population from learning the substance of the test as such.

DISCUSSION

METHOD OF APPROACH.

The technical method of approach in developing the CPM test is to design and try out several sets of traffic samples for use in air traffic control simulation in order to form a standardized testing instrument. This involves the working out of a set of measures which can be used in normative distributions.

Figure 1 shows the test environment with two controllers working in the NAFEC dynamic air traffic control simulator. The controllers worked the same sector and handled the identical sample of traffic, which was separately fed to them by the simulator. They worked without assistant controllers so that all results would be attributable to them as individuals. The traffic was generated by a large-scale digital simulator and directed by simulator operators who represented pilots in the real air traffic control (ATC) system. The "pilots" and the controllers communicated over simulated radio frequencies. In this particular experiment, a broadband system with shrimpboat tracking was simulated (see figure 1).

The computer recorded aircraft events which were reflective of the safe and expeditious movement of air traffic. At the end of an hour, the computer printed out a summary of performance measure scores based on aircraft events. The performance measures used are listed and defined later in this chapter. In addition to the performance measures, heart rate was taken during every run.

HYPOTHESIS.

The hypothesis of experiment I was that it would be possible to build equivalent forms of the traffic sample test by relying on the interaction of sector-structure complexity and traffic density level. What it was believed might occur can be best explained through use of figure 2. In this figure, it can be seen that there might be combinations of the level of traffic (in terms, say, of number of aircraft to be serviced per hour), and the geographic complexity of the sector (conceptually, the number of routes to be watched, the number of intersections involved, and the geographic size) which might appear quite different, but would yield the same average level of score and thus represent different tests of equivalent difficulty. The design of experiment I, then, was based on the concept illustrated in figure 2, except that two, not three, sector structures were used.

PROCEDURE.

For this pilot study, six qualified enroute air traffic controllers from the NAFEC evaluation group served as subjects. Every subject worked in every sector/traffic-level combination condition, of which there were six. Two sector structures were chosen so as to represent broad differences in normal sector structures. These sectors were chosen from a large library of sectors available at NAFEC from a previous project which had had contact with many

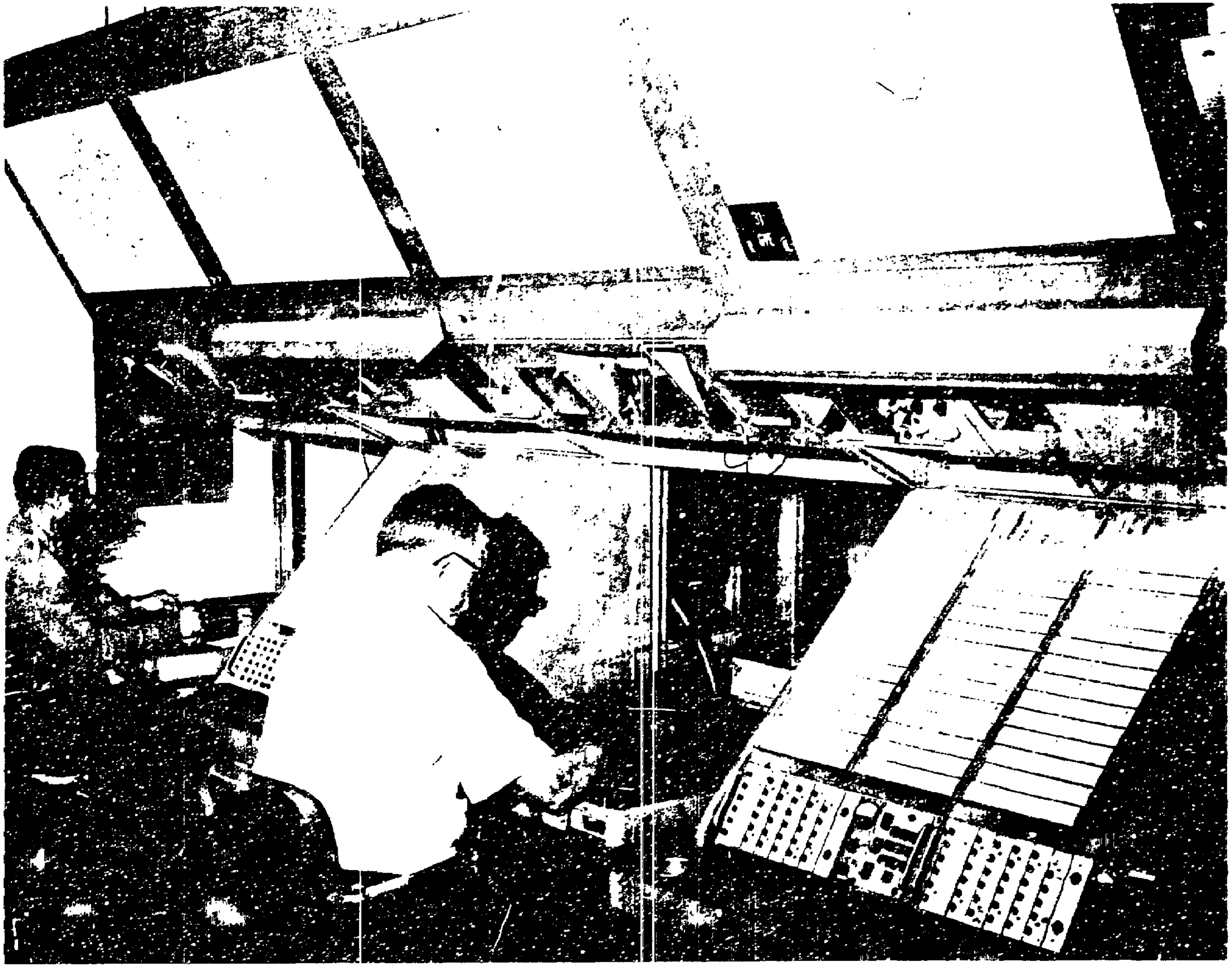
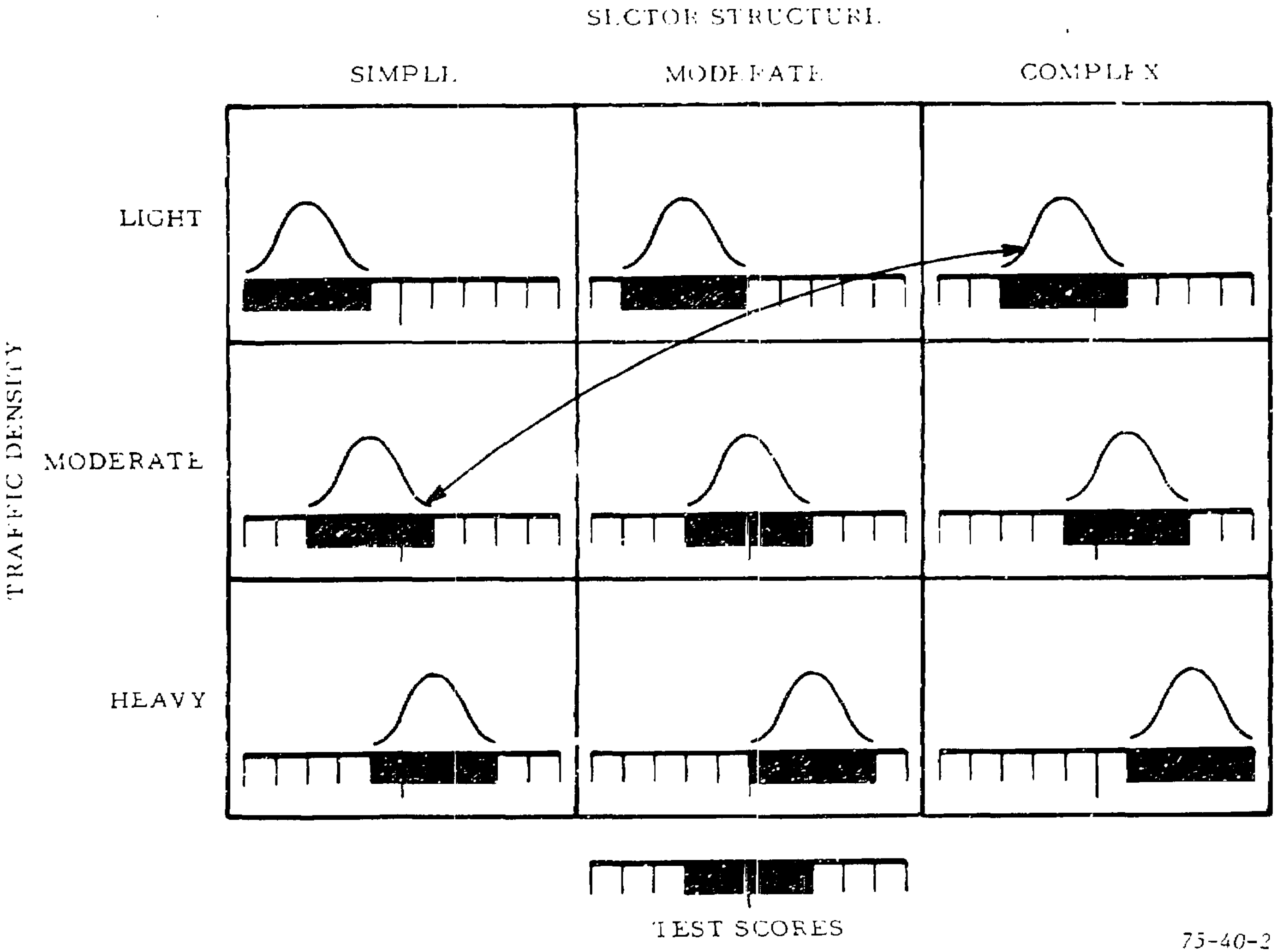


FIGURE 1. PHOTOGRAPH OF SUBJECTS DURING TEST SESSION



75-40-2

FIGURE 2. PICTORIAL STATEMENT OF SECTOR/DENSITY INTERACTION HYPOTHESIS

sectors from all over the country. The sector maps appear in figure 3. Three traffic density levels were chosen, which are describable either as 40, 50, and 60 aircraft to be handled per hour, or, as 8 aircraft present at all times, 10 aircraft present at all times, and 12 aircraft present at all times (in approximate terms).

The experimental design is presented in table 1. It is definable as a split plot factorial p.qr type design in the terminology of Kirk (reference 4, p. 300). The six subjects were divided randomly into two groups of three each so as to provide a control for the time order in which they would work the two different sectors. Group 1 worked sector 14 first, then sector 16. Group 2 worked sector 16 first, then sector 14. The order of encountering the three densities was counterbalanced, as may be seen in the table, in that the letters a through f represent the order in which each subject encountered the six conditions.

The experimental sessions were 1 hour and 15 minutes long; 15 minutes for warmup and 1 hour during which data were taken.

MEASURES.

Two types of measures were used: performance measures, which were made up of various data elements; and the heart rate measure.

PERFORMANCE MEASURES.

Data Elements. Eight basic performance data elements were combined with 2 traffic sample parameters to make a set of 10 performance measures. The combinations were such as to create more meaningful measures. Generally, the effect was to convert the measure to a proportion of possible outcomes of a given type.

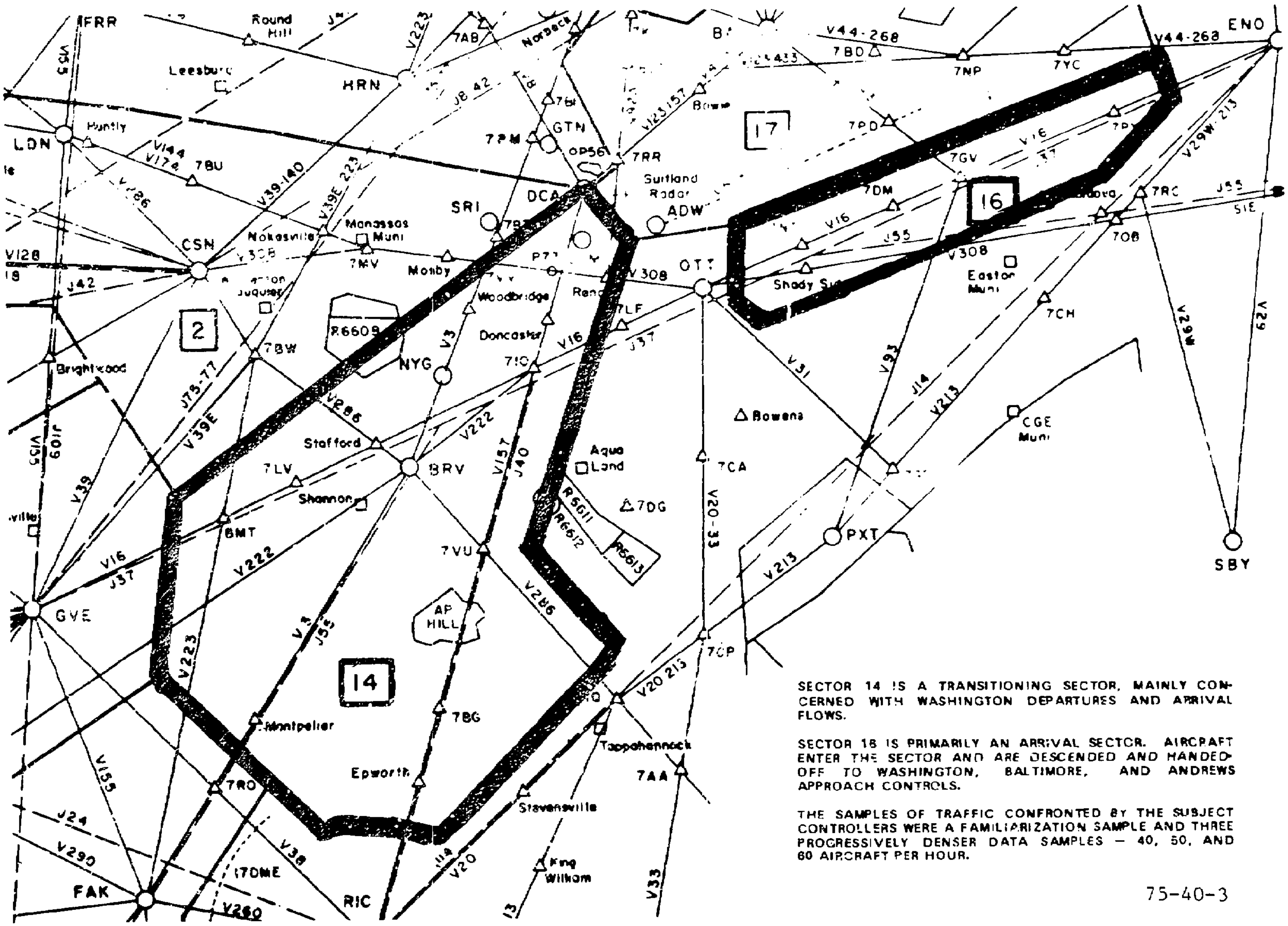
The basic data elements are defined as follows:

1. Number of Conflictions. Conflictions were violations of the separation standard, which was in this instance, "less than 4.50 nautical miles (nmi) and 950 feet." The computer recorded and counted these.

2. Number of Delays. The computer counted the number of delays to aircraft in the following manner:

a. Start time delays. These delays were of aircraft not allowed to begin their flight at their scheduled start time. A 90-second "fudge" factor was provided in each instance to cover delay by the simulated adjacent sector controller and insure that this did not impinge on the subject controller's score.

b. Hold delays. These delays were of aircraft flying in the system airspace which were given a hold message by the test controller. They entered the classical "racetrack" holding pattern.



SECTOR 14 IS A TRANSITIONING SECTOR, MAINLY CONCERNED WITH WASHINGTON DEPARTURES AND ARRIVAL FLOWS.

SECTOR 16 IS PRIMARILY AN ARRIVAL SECTOR. AIRCRAFT ENTER THE SECTOR AND ARE DESCENDED AND HANDED-OFF TO WASHINGTON, BALTIMORE, AND ANDREWS APPROACH CONTROLS.

THE SAMPLES OF TRAFFIC CONFRONTED BY THE SUBJECT CONTROLLERS WERE A FAMILIARIZATION SAMPLE AND THREE PROGRESSIVELY DENSER DATA SAMPLES — 40, 50, AND 60 AIRCRAFT PER HOUR.

75-40-3

FIGURE 3. SECTOR MAPS

TABLE 1. EXPERIMENTAL DESIGN

Traffic Density	Group	Sector 14		Sector 16	
		Subject	*	Subject	*
40 flights per hour	1	1	a	1	d
		2	b	2	e
		3	c	3	f
	2	4	d	4	a
		5	e	5	b
		6	f	6	c
50 flights per hour	1	1	b	1	e
		2	c	2	f
		3	a	3	d
	2	4	e	4	b
		5	f	5	c
		6	d	6	a
60 flights per hour	1	1	c	1	f
		2	a	2	d
		3	b	3	e
	2	4	f	4	c
		5	d	5	a
		6	e	6	b

*The letters a through f represent the order in which each subject encountered the different conditions.

c. Turn delays. These delays were recorded whenever an aircraft was given a heading change, the intent of which was to make more room or "stretch" the path of the subject aircraft. It provided for "Make a 360...." type delays. In order to allow for normal turning along an airway or leaving a holding pattern, the turn had to be greater than 100 seconds in duration, or approximately 300°, to be counted as a delay.

3. Cumulative Delay Time. This was the sum of the duration of all of the events described above (2 a, b, and c) (Delays) expressed in seconds.

4. Number of Completed Flights. This was the total of controlled flights which were changed from the active frequency to a handoff frequency. Thus, the number of aircraft which transited the sector to a position of "completion" was recorded.

5. Number of Air/Ground Contacts. This was the total number of messages initiated by the subject controller.

6. Cumulative Air/Ground Communications Time. This was the duration in seconds of all of the subject's messages to controlled aircraft.

7. Number of Aircraft Handled. This was the sum of all controlled aircraft confronted and accepted by the subject in the hour-long sample. This included those aircraft which had entered the sector and had not transited to points of completion.

8. Idents. This was the number of times the pilot was requested by the controller to verify his identity by beacon.

9. Number of Aircraft in the Sample. This was the total number of aircraft in the traffic sample. It differs from 7 in that the subject may not have accepted all of the aircraft handed off to him from the adjacent sector in the sample.

10. Number of Completable Flights. This was the number of flights, determined beforehand, which could reasonably be expected to reach their destinations or be handed off before the data hour ended.

Performance Measures. The performance measures are combinations of the above data elements. The elements are placed into ratios, or other combinations or permutations for more meaningful measurement. For example. Measure 1 is obtained by dividing Data Element 1 by Data Element 7. (For a discussion of this point, see reference 1.) The measures are defined as follows:

1. Number of Conflictions/Number of Aircraft Handled.

$$\frac{\text{Data Element 1}}{\text{Data Element 7}}$$

2. Number of Conflictions/Number of Delays.

$$\frac{\text{Data Element 1}}{\text{Data Element 2}}$$

3. Number of Delays/Number of Aircraft in Sample.

$$\frac{\text{Data Element 2}}{\text{Data Element 9}}$$

4. Cumulative Delay Time/Number of Aircraft in Sample.

$$\frac{\text{Data Element 3}}{\text{Data Element 9}}$$

5. Number of Completed Flights/Number of Completable Flights.

$$\frac{\text{Data Element 4}}{\text{Data Element 10}}$$

6. Number of Contacts/Number of Aircraft Handled.

$$\frac{\text{Data Element 5}}{\text{Data Element 7}}$$

7. Communication Time/Number of Contacts.

$$\frac{\text{Data Element 6}}{\text{Data Element 5}}$$

8. Number of Aircraft Handled/Number of Aircraft in Sample.

$$\frac{\text{Data Element 7}}{\text{Data Element 9}}$$

9. Correlation Hold-Delay Transformation.

This is the product-moment correlation coefficient computed on the basis of data points every 10 minutes within the data hour using Data Elements 3 and 7 and transformed using the z transform.

10. Surplus Idents.

Data Element 8 minus Data Element 9.

HEART RATE MEASURE. The above are the performance measures. Another measurement taken was the heart rate of the controllers while working the traffic problems in the simulator. Heart rate was measured for each subject during each run, and the heart rate measure was also subjected to the analysis of variance. Heart rate is well accepted as a measure of effort, at least of physical effort, and to some extent of generalized effort and pressure. Heart rate is elevated over its normal resting rate in pressure situations. It was of interest here as a measure of workload.

The procedure used was the taking of a resting heart rate before the actual experimental run, and then the monitoring of the heart rate during the hour-long run. The heart rate for the hour run was divided by the number of minutes the run lasted (60) to get the average heart rate during the run. Then the difference (presumably the amount of elevation) between the heart rate at rest and the heart rate at work with the particular traffic sample/sector situation was computed and used as one piece of data concerning the run.

RESULTS

PERFORMANCE DATA.

GENERAL. A simplified experimental design is shown in table 2. The basic data for each subject, which will later be discussed statistically, can be seen in histogram form in figure 4 for each of the 10 performance measures. The sector/density combination means and standard deviations are also given.

In general, the results indicate that the hypothesis of interaction between sector and density in affecting performance was not sustained. There was little difference shown in the measures between the two sectors. Great difference was shown between the three levels of traffic. It appears that construction of sector structure/density combinations is not available, or necessary, as a route to the goal of comparably difficult traffic problems, but rather that the use of comparable traffic density levels with almost any representative sector structure would be adequate to the purpose. This information will serve to guide future steps in the process of criterion development but will, of course, come under review and validation as the process continues. It should be pointed out that this finding does not deny differences among field traffic control sectors; they differ in both traffic density and structure, simultaneously and irregularly. The two factors were varied independently and regularly in this experiment.

STATISTICAL TREATMENT. The basic experimental design was discussed earlier. The role of this experiment as a probe in a larger pursuit, rather than as an end in itself, explains the small number of subjects and number of runs under the various conditions. Within these limitations, the analysis of variance was performed on the measures. Two analyses were done. In the first, the original design, a "groups" factor based on the order or sequence in which the subjects encountered the two sectors, was included. In the second, after

TABLE 2. EXPERIMENTAL DESIGN (SIMPLIFIED)

Traffic Level (Flights Per Hour)	Sector 14 Controller No.	Sector 16 Controller No.
40	1	1
	2	2
	3	3
	4	4
	5	5
	6	6
50	1	1
	2	2
	3	3
	4	4
	5	5
	6	6
60	1	1
	2	2
	3	3
	4	4
	5	5
	6	6
TOTAL		36

examining the results of the first analysis, the groups factor was omitted because the impact of the groups factor appeared to be diverse and slight. The second design, then, was a three-factor analysis involving the variables of subjects (6), sectors (2), and traffic densities (3), in which every subject worked in every condition (table 2). The results will be discussed in terms of this design. It should be remembered that if the assumptions of this design were to be violated, the outcome would be in the direction of finding a higher frequency of statistically significant outcomes, not a lesser one (reference 5).

The results of the second analysis of variance were followed up more closely as to the differences between sectors at a given density by use of a nonparametric test of bivariate symmetry developed by Hollander (reference 6). This test was done because there were a few sector/density interactions, but more importantly, because it was noticed that the standard deviations sometimes changed. This test checked the distribution similarity in all respects between the two sectors, including both central tendency and variation. The test is quite laborious since it involves an exact computation of probabilities. It is intended for use with small sample sizes and it works in two stages; it

FIGURE 4. HISTOGRAMS OF PERFORMANCE DATA[1]

Figure	
4a	Number of Conflictions/Number of Aircraft Handled
4b	Number of Conflictions/Number of Delays
4c	Number of Delays/Number of Aircraft in Sample
4d	Cumulative Delay Time/Number of Aircraft in Sample
4e	Number of Completed Flights/Number of Completable Flights
4f	Number of Contacts/Number of Aircraft Handled
4g	Communication Time/Number of Contacts
4h	Number of Aircraft Handled/Number of Aircraft in Sample
4i	Correlation Hold-Delay Transformation
4j	Surplus Idents

[1] In these histograms, the numerical interval designations represent the lower limits of the intervals.

INTERVAL	SR 14 D1	SR 14 D2	SR 14 D3	SR 16 D1	SR 16 D2	SR16 D3
.220)						
.200)	*	*	*			
.180)		*			*	*
.160)			**		*	
.140)						*
.120)	*	*			*	*
.100)	**	*		*		*
.080)					*	
.060)		*	*	***	*	*
.040)		*	*		*	*
.020)	*			**		
.000)	*		*			
MEAN	.090	.123	.108	.058	.110	.112
S DEV	.072	.067	.086	.032	.059	.050
N	6.	6.	6.	6.	6.	6.

*ALL GROUPS COMBINED

MEAN	.1003
S DEV	.0623
MAXIMUM	.2100
MINIMUM	.0000

75-40-4a

FIGURE 4a. NUMBER OF CONFLICTIONS/NUMBER OF AIRCRAFT HANDLED

```
 SR 14 D1 | SR 14 D2 | SR 14 D3 | SR 16 D1 | SR 16 D2 | SR16 D3

INTERVAL

 4.400 )
 4.000 )**
 3.600 )
 3.200 )
 2.800 )
 2.400 )*
 2.000 )
 1.600 )                                        *
 1.200 )*
  .800 )                                       *
  .400 )           *          *         ***              *
  .000 )**       *****      *****       ***    *****    *****

MEAN       1.987      .313       .165      .353      .417      .236
S DEV      1.844      .228       .168      .309      .583      .215
N          6.         6.         6.        6.        6.        6.
*ALL GROUPS COMBINED

          MEAN       .5785
          S DEV      .9904
          MAXIMUM   4.0000
          MINIMUM    .0000
```

75-40-4b

FIGURE 4b. NUMBER OF CONFLICTIONS/NUMBER OF DELAYS

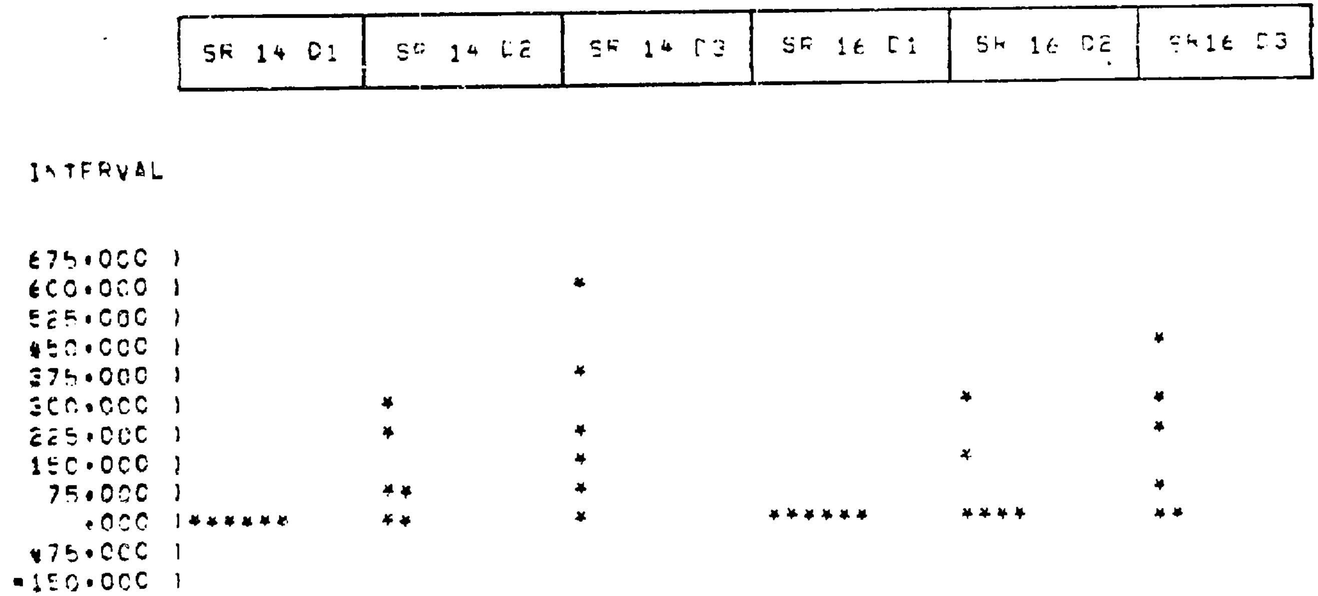
INTERVAL	SR 14 D1	SR 14 C2	SR 14 C3	SR 16 C1	SR 16 D2	SR16 C3
.900						
.800			*			*
.700		*	*		*	**
.600		*	*		*	*
.500		*	*		*	
.400		*	*			
.300		*	*			*
.200				*	**	*
.100	*	*		**	*	
.000	*****			***		
-.100						
-.200						
MEAN	.043	.457	.605	.095	.427	.582
S DEV	.037	.203	.192	.100	.267	.235
N	6.	6.	6.	6.	6.	6.

*ALL GROUPS COMBINED

MEAN	.3681
S DEV	.2848
MAXIMUM	.8700
MINIMUM	.0000

75-40-4c

FIGURE 4c. NUMBER OF DELAYS/NUMBER OF AIRCRAFT IN SAMPLE



	SR 14 D1	SR 14 D2	SR 14 D3	SR 16 D1	SR 16 D2	SR16 D3
MEAN	5.152	144.903	274.106	3.632	104.567	201.778
S DEV	10.363	110.073	212.614	4.171	135.178	150.057
N	6.	6.	6.	6.	6.	6.

*ALL GROUPS COMBINED

MEAN	122.3562
S DEV	159.3366
MAXIMUM	612.0300
MINIMUM	.0000

75-40-4d

FIGURE 4d. CUMULATIVE DELAY TIME/NUMBER OF AIRCRAFT IN SAMPLE

INTERVAL	SR 14 D1	SR 14 D2	SR 14 D3	SR 16 D1	SR 16 D2	SR16 D3
1.000)	*			****		
.950)	***	*		*		*
.900)	*	*		*	**	
.850)	*		**			*
.800)		*			**	*
.750)		**	**		*	
.700)		*	*			**
.650)					*	
.600)						*
.550)						
.500)			*			
.450)						
MEAN	.947	.83[illegible]	.752	.982	.817	.790
S DEV	.046	.098	.133	.033	.036	.116
N	6.	6.	6.	6.	6.	6.

*ALL GROUPS COMBINED

MEAN	.852[illegible]
S DEV	.121[illegible]
MAXIMUM	1.0000
MINIMUM	.5200

75-40-4e

FIGURE 4e. NUMBER OF COMPLETED FLIGHTS/NUMBER OF COMPLETABLE FLIGHTS

INTERVAL	SR 14 D1	SR 14 D2	SR 14 D3	SR 16 D1	SR 16 D2	SR16 D3
8.000)						
7.600)		*				
7.200)						
6.800)	*					
6.400)			*			
6.000)	*	*			*	
5.600)	**	*		***	**	*
5.200)	*	**	**	**	*	**
4.800)	*	*	***		*	**
4.400)				*	*	*
4.000)						
3.600)						
MEAN	5.775	5.901	5.407	5.425	5.492	5.123
S DEV	.690	.925	.699	.458	.567	.433
N	6.	6.	6.	6.	6.	6.

*ALL GROUPS COMBINED

MEAN	5.5187
S DEV	.6553
MAXIMUM	7.6200
MINIMUM	4.4500

75-40-4f

FIGURE 4f. NUMBER OF CONTACTS/NUMBER OF AIRCRAFT HANDLED

INTERVAL	SR 14 D1	SR 14 D2	SR 14 D3	SR 16 D1	SR 16 D2	SR16 D3
4.400						
4.200					*	
4.000						
3.800						
3.600	*		*		*	*
3.400	**	**		***		**
3.200	*				**	
3.000		**	***	**	*	*
2.800	**	*		*	*	**
2.600		*	**			
2.400						
2.200						
MEAN	3.325	3.161	3.060	3.235	3.447	3.232
S DEV	.332	.337	.378	.264	.43[illegible]	.325
N	6.	6.	6.	6.	6.	6.

*ALL GROUPS COMBINED

MEAN	3.2433
S DEV	.3577
MAXIMUM	4.2900
MINIMUM	2.6700

75-40-4g

FIGURE 4g. COMMUNICATION TIME/NUMBER OF CONTACTS

INTERVAL	SR 14 D1	SR 14 D2	SR 14 D3	SR 16 D1	SR 16 D2	SR16 D3
1.040)						
1.000)	******	**		******		*
.960)			*			*
.920)		**	*		****	*
.880)			*			
.840)		**				*
.800)			*			
.760)			*		**	*
.720)						*
.680)						
.640)			*			
.600)						
MEAN	1.000	.93[illegible]	.842	1.000	.880	.883
S DEV	.00[illegible]	.063	.119	.000	.078	.113
N	6.	6.	6.	6.	6.	6.

*ALL GROUPS COMBINED

MEAN	.9231
S DEV	.0956
MAXIMUM	1.0200
MINIMUM	.6500

75-40-4h

FIGURE 4h. NUMBER OF AIRCRAFT HANDLED/NUMBER OF AIRCRAFT IN SAMPLE

INTERVAL	SR 14 D1	SR 14 D2	SR 14 D3	SR 16 D1	SR 16 D2	SR16 D3
.750)						
.600)						
.450)				*	*	**
.300)	*	*	*	*	*	
.150)	**	**			*	*
.000)	**	*	*	****	*	*
-.150)	*	*	*		**	
-.300)		*	**			*
-.450)			*			
-.600)						
-.750)						*
-.900)						
MEAN	.143	.12[illegible]	-.049	.174	.2[illegible]4	.076
S DEV	.158	.22[illegible]	.274	.231	.271	.43[illegible]
N	6.	6.	6.	6.	6.	6.

*ALL GROUPS COMBINED

MEAN	.1112
S DEV	.2713
MAXIMUM	.5860
MINIMUM	-.6620

75-40-41

FIGURE 4i. CORRELATION HOLD-DELAY TRANSFORMATION

INTERVAL	SR 14 D1	SR 14 D2	SR 14 D3	SR 16 D1	SR 16 D2	SR16 D3
67.500)						
60.000)						
52.500)						*
45.000)					*	
37.500)						
30.000)						*
22.500)					*	
15.000)		**	*		*	*
7.500)	*	**	***		*	**
.000)	*****	**	**	******	**	*
-7.500)						
-15.000)						
MEAN	4.833	9.333	8.500	.833	16.667	22.000
S DEV	3.971	6.250	5.206	1.329	17.478	19.256
N	6.	6.	6.	6.	6.	6.

-ALL GROUPS COMBINED

MEAN	10.3611
S DEV	12.6562
MAXIMUM	56.0000
MINIMUM	.0000

75-40-4j

FIGURE 4j. SURPLUS IDENTS

gives a result indicating acceptance or rejection of the hypothesis, or, if unable to do that, it gives a random decision value (L). This can roughly be interpreted as a less certain statement about the hypothesis. It is the probability of rejecting the hypothesis in a randomized decision. If this value is low (e.g., .10), it would appear safe, but less than certain, to accept the hypothesis of equality. The L value did occur in a few instances, as will be discussed later.

INTERPRETATION OF DATA. The data of the experiment were the measures of performance obtained by the six subjects under the six conditions of the experiment. As mentioned earlier, the basic data for each subject and the means and standard deviations for each of the six conditions appear in figure 4. Table 3 summarizes the results of the analysis of variance. Table 4 summarizes the results of the test of the similarity in all respects of the distributions of scores obtained in each sector at each density.

The 10 performance measures will now be discussed in order.

MEASURE 1--NUMBER OF CONFLICTIONS/NUMBER OF AIRCRAFT HANDLED. This ratio could be interpreted as the rate of conflictions per aircraft handled. Since the number of aircraft handled increased with the scheduled traffic densities, as did the conflictions, it is not surprising that this ratio remained constant (or more or less so) across the three densities. It was also similar for the two sectors. There were no statistically significant differences with sector or density, nor was the interaction significant. It should be pointed out, parenthetically, that any number of conflictions scored here does not mean that the real system has that level of conflictions; the system is safe. The traffic densities handled here are considerably higher than those in the real system, and they are handled here by one man rather than a team of men.

MEASURE 2--NUMBER OF CONFLICTS/NUMBER OF DELAYS. This measure represents an attempt to encapsulate the comparative tendency of various controllers to err, if they are going to err, in the direction of delays rather than conflictions, or vice versa.

For this measure, it is believed that the sector/density interaction indicated in Table 3 is simply spurious. It can be seen from the mean values presented in the histogram that there were a few odd values in two of the conditions which strongly affected the means. The density effect indicated by the analysis of variance also seems irregular and probably spurious. The bivariate test indicated no statistically significant difference between the sectors at the various respective densities.

MEASURE 3--NUMBER OF DELAYS/NUMBER OF AIRCRAFT IN SAMPLE. This ratio might have been expected to remain constant, or at least similar, across densities. It generally represents the number of delayed aircraft out of those in the sample "available," as it were, for delay. Apparently the number of delays increased faster than the number of aircraft in the three traffic samples did. There was, then, a firm density effect, but no sector effect or interaction.

TABLE 3. RESULTS OF ANALYSIS OF VARIANCE

	Sector	Density	Sector By Density
1. Number of Conflictions/Number of Aircraft Handled			
2. Number of Conflictions/Number of Delays		S*	S
3. Number of Delays/Number of Aircraft in Sample		S	-**
4. Cumulative Delay Time/Number of Aircraft in Sample		S	-
5. Number of Completed Flights/Number of Completable Flights		S	-
6. Number of Contacts/Number of Aircraft Handled		S	-
7. Communication Time/Number of Contacts		-	S
8. Number of Aircraft Handled/Number of Aircraft in Sample		S	S
9. Correlation Hold-Delay Transformation		-	-
10. Surplus Idents		S	S

* S = Significant at the .05 level
** - = Not significant at the .05 level

TABLE 4. RESULTS OF TEST OF BIVARIATE SYMMETRY OF SECTORS

	Density 1	Density 2	Density 3
1. Number of Conflictions/Number of Aircraft Handled	NS*	NS	NS
2. Number of Conflictions/Number of Delays	NS	NS	NS
3. Number of Delays/Number of Aircraft in Sample	NS	NS	NS
4. Cumulative Delay Time/Number of Aircraft in Sample	NS	NS	(L=.10)**
5. Number of Completed Flights/Number of Completable Flights	NS	NS	NS
6. Number of Contacts/Number of Aircraft Handled	NS	NS	NS
7. Communication Time/Number of Contacts	NS	NS	NS
8. Number of Aircraft Handled/Number of Aircraft in Sample	NS	NS	(L=.10)
9. Correlation Hold-Delay Transformation	NS	NS	NS
10. Surplus Idents	(L=.80)	NS	(L=.80)

* NS = not significant at the .05 level

** L = The L probability is computed for cases in which the test statistic is equal to the critical value. It is the probability associated with randomly rejecting the null hypothesis. When this probability is close to zero, it is proper to accept the null hypothesis (i.e., no sector difference).

MEASURE 4--CUMULATIVE DELAY TIME/NUMBER OF AIRCRAFT IN SAMPLE. The total cumulative delay time divided by the number of aircraft in the sample results in the average delay time (in seconds) per aircraft in the sample. It will be remembered that this delay time includes delay for handoffs into the subject's sector and enroute delays. The delay time differences with density were regular and significant. There was no sector main effect, nor was there a significant interaction. The bivariate test picked up an L value between sectors at the highest density, but the value was low, and so it can be considered that there was no significant sector effect.

MEASURE 5--NUMBER OF COMPLETED FLIGHTS/NUMBER OF COMPLETABLE FLIGHTS. The ratio behaved very regularly. There was a significant change with density, a wide individual controller variation, and no difference as a function of sector structure.

MEASURE 6--NUMBER OF CONTACTS/NUMBER OF AIRCRAFT HANDLED This is the number of contacts required per aircraft handled; i.e., accepted and moved through the sector. There were about five to seven contacts per aircraft. There was a statistically significant difference with density, but very probably not a meaningful one. The spread among subjects was narrow and may not be very meaningful. This measure may need to be dropped or modified.

MEASURE 7--COMMUNICATION TIME/NUMBER OF CONTACTS. This is the average time spent in talking each time there was communication between the controller and pilot. A tendency to decrease with the number of aircraft (traffic density) being faced is noticeable. There was some irregularity to be noted, however, in the means for the six conditions, and this resulted in a statistically significant interaction in the analysis of variance. Very likely, however, this was exactly that, an irregularity, and not a meaningful interaction. There were no significant differences found between the two sector distributions at corresponding densities. Individual differences in being able to adapt communication length to situational demands are probably important.

MEASURE 8--NUMBER OF AIRCRAFT HANDLED/NUMBER OF AIRCRAFT IN SAMPLE. In the lowest traffic density, all subjects handled 100 percent of the aircraft in both sectors. At the middle density, the mean values were 93 percent for sector 14 and 88 percent for sector 16; a 5-percent difference favoring sector 14. But at the highest density, the mean values were 84 percent for sector 14 and 88 percent for sector 16; a 4-percent difference, this time favoring sector 16. For this reason, the analysis of variance indicated a statistically significant interaction between density and sector in addition to the normally significant main effect for density. Responding to the interaction and looking at the densities separately, we see that at the lowest density there was no difference at all in the distribution; i.e., everyone handled all the aircraft. The nonparametric test found essentially that the distributions at the middle and high densities were not significantly different, despite the 4- or 5-percent differences mentioned above. In short, there does not seem to be a clear-cut conclusion possible in regard to the indications of this particular measure in this instance.

MEASURE 9--CORRELATION HOLD-DELAY TRANSFORMATION. This measure was included in this experiment as a result of some observations in previous work (reference 1). There the correlation between the number of delays (or delay time) in a run and the number of aircraft handled in the same run seemed to be a measure which was, in itself, sensitive to changes in density and controller ability (as indicated on other grounds). For the measure here, successive 10-minute periods of the run were used as the unit and a correlation was computed for each run from these within-run data, even though it was realized that successive time periods of the same run do not represent statistically independent data points. The measure used is the Z transformation of the correlations for computation purposes.

The measure did vary with density. The variation with density was not statistically significant (the probability value was .16, not .05 or less), but the trend was regular with density and in the direction predicted by the earlier work which was referred to above (reference 1); i.e., a decreasing correlation, tending toward a negative correlation as traffic density increased and decreasing as individual proficiency was reflected as lower on other measures.

This measure also indicated, although the indication was not at all close to being statistically significant, that perhaps there was a slight tendency for sector 16 to be easier.

MEASURE 10--SURPLUS OF IDENTS OVER NUMBER OF AIRCRAFT HANDLED. An "ident" is shorthand for getting identification from an aircraft by means of a request to the pilot to activate certain beacon equipment. This is done once, in broadband (raw radar) control, upon acceptance of the handoff. On subsequent occasions, the procedure is resorted to if doubt about the identity of any aircraft being tracked arises. Therefore, the number of idents resorted to above the number accepted (i.e., handled) was computed as a difference.

The statistical analysis of variance indicated a significant difference with density but also a sector-by-density interaction. The interaction was so complex as to suggest that part of it at least might be due to chance fluctuations despite the statistical result. The number of surplus idents at low density was higher for sector 14 than for sector 16, but was higher for sector 16 than for sector 14 at both of the higher densities. This would seem to indicate that a special situation involving some extra shrimpboat handling and identification difficulty was present in sector 16, as was confirmed subjectively.

REVIEW. The hypothesis stated that it was expected there would be such a strong interaction between sector and density that equivalent distributions might result from combinations of sector and density. In general, this strength of interaction did not result. On the contrary, the effect of sector structure was generally negligible, whereas the effect of density was most often very strong. It would appear, in short, that all that is required for parallel forms is to have the same level of traffic density, without regard to sector structure.

It should be remembered that the independent variation of structure of sectors and traffic density is not possible in the field, which is why this finding may seem to contradict field experience.

HEART RATE DATA.

The histograms in figure 5 show the basic heart rate difference data. To review, for each run by each subject, a subtraction was made between his average heart rate per minute during the run and his resting heart rate that day, such as to indicate the increase the run made over the rest rate. The analysis of variance indicates a significant main effect between sectors and between densities, and no significant interaction. The mean scores are plotted in figure 6. The bivariate symmetry test indicates that only at the lowest density are the distributions different between the sectors (this includes the mean and standard deviation). The difference as a function of density was expected, but the difference as a function of sector was surprising in view of the previous analyses. The loss, due to technical difficulties with the data for 3 of the 36 runs, might have some bearing on the matter. Also to be considered was the fact that the differences between the two sectors at the different densities may not have been very great in absolute terms. The differences were approximately 13, 9, and 5 beats per minute between the means for sectors 14 and 16 at the low, medium, and high densities, respectively, with the sector 16 values always higher.

Nonetheless, there would seem to be some indication here that more effort was required when working sector 16. While it was not a resounding difference or even very conclusive, it would seem wise to consider the possibility that the two sectors might have required different levels of effort to produce the same average performance.

SECTOR CHARACTERISTICS AND PERFORMANCE.

The indication that the sectors were essentially similar, despite having been chosen on the basis of being apparently quite different, was surprising. Collaboration was therefore sought by reference to an important recent theoretical analysis of air traffic procedures and movements. This is the work by Ratner et al., of Stanford Research Institute (SRI). SRI has developed what it feels is a mathematical expression which is reflective of the difficulty of a sector. It is based, among other things, on the number of intersections in a sector.

In that respect, at least, the two sectors used here are remarkably different, since one sector has only one major intersection and the other has several. Using a nomograph prepared by SRI (reference 7) and the equations described in an associated report (reference 8), data from an average run were examined and the parameters required by the formulations were derived.

Using the derived parameters, the Stanford CDI (Control Difficulty Index) was computed for the six sector/density combinations. Higher CDI values were found for sector 14 than for sector 16. The CDI data are plotted in figure 7. On the assumption that number of delays was an index of actual control difficulty,

INTERVAL	SR 14 D1	SR 14 C2	SR 14 C3	SR 16 C1	SR 16 D2	SR16 C3
40.000)						
35.000)					*	
30.000)				*		*
25.000)						
20.000)		*	*		*	**
15.000)			***	*	*	*
10.000)	**	**	*	*	*	*
5.000)				**		*
.000)	**	*	*		*	
-5.000)	**	*				
-10.000)						
-15.000)						
MEAN	3.417	9.500	14.833	16.000	12.200	19.333
S DEV	7.499	9.657	7.731	9.055	13.424	9.459
N	6.	5.	6.	5.	5.	6.

-ALL GROUPS COMBINED

MEAN	13.4545
S DEV	10.4518
MAXIMUM	39.0000
MINIMUM	-4.5000

75-40-5

FIGURE 5. HISTOGRAMS OF HEART RATE DATA

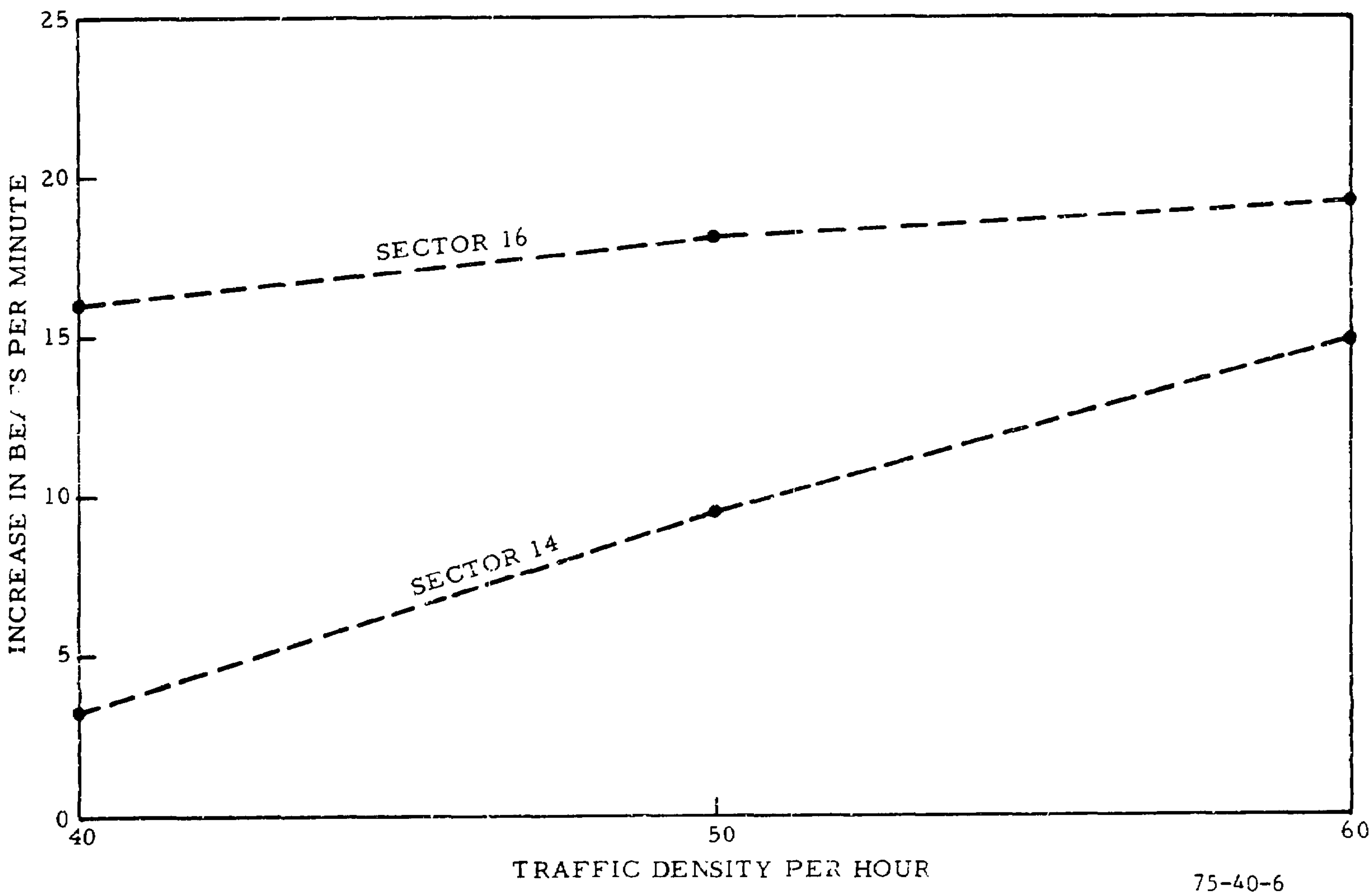


FIGURE 6. AVERAGES OF HEART RATE DATA

the average delay score was also plotted in figure 7. In order to plot both scores on the same scale, they were each expressed as a proportion of their own highest value. It may be seen from the figure that there was some agreement but also some difference between the empirical data and the mathematically derived index values.

Realizing the limitations of the procedure described above, the main point is only that such mathematical approximations can be validated and probably refined by the use of real time simulation. Occasional attempts to apply such models and verify them will be one part of the current project, since a method of determining, at least approximately, the relative difficulty of a traffic sample/sector combination in advance of any runs would be a useful tool for this work.

DIGRESSION.

After this long discussion of the difference between sectors, a digression would appear desirable to restore the focus to the basic purpose of the work, which is, after all, not the difference between sectors, but the difference between individuals. For this reason, the score profiles on selected measures for two subjects on the two sectors (at the middle density) are presented in figure 8. These subjects were chosen, for illustrative purposes, to be those whose profiles on the basic measures differed the most. The profiles are in terms of standard scores, which are a method of reducing scores to common units. (For further information, the reader is referred to standard psychometric statistics sources, such as McNemar, reference 9).

It may be seen from the profiles that the two controllers perform quite differently, and that the examination of such profiles could be diagnostically informative. Looking at the top half of the illustration, we see the performance profiles of the two controllers when working with sector 14. Controller A has his lower scores on the left half of the profile; controller B has his lower scores on the right half of the profile. Looking at the lower half of the page, it can be seen that the pattern is essentially repeated: controller A has his lower scores on the left half of the profile and controller B has his lower scores on the right half of his profile. The two controllers followed their same patterns of action in both sectors. It happens, incidentally, that the three scores on the left side of the profiles are of a negative sort; high scores mean more conflictions, more delays, and more delay time. On the right half of the profile, the scores are more positive; more completed flights, more of the available aircraft handled, and a more positive score on the correlation-transformation index.

This illustration is intended to show how such profiles can be instructive concerning individual performance patterns.

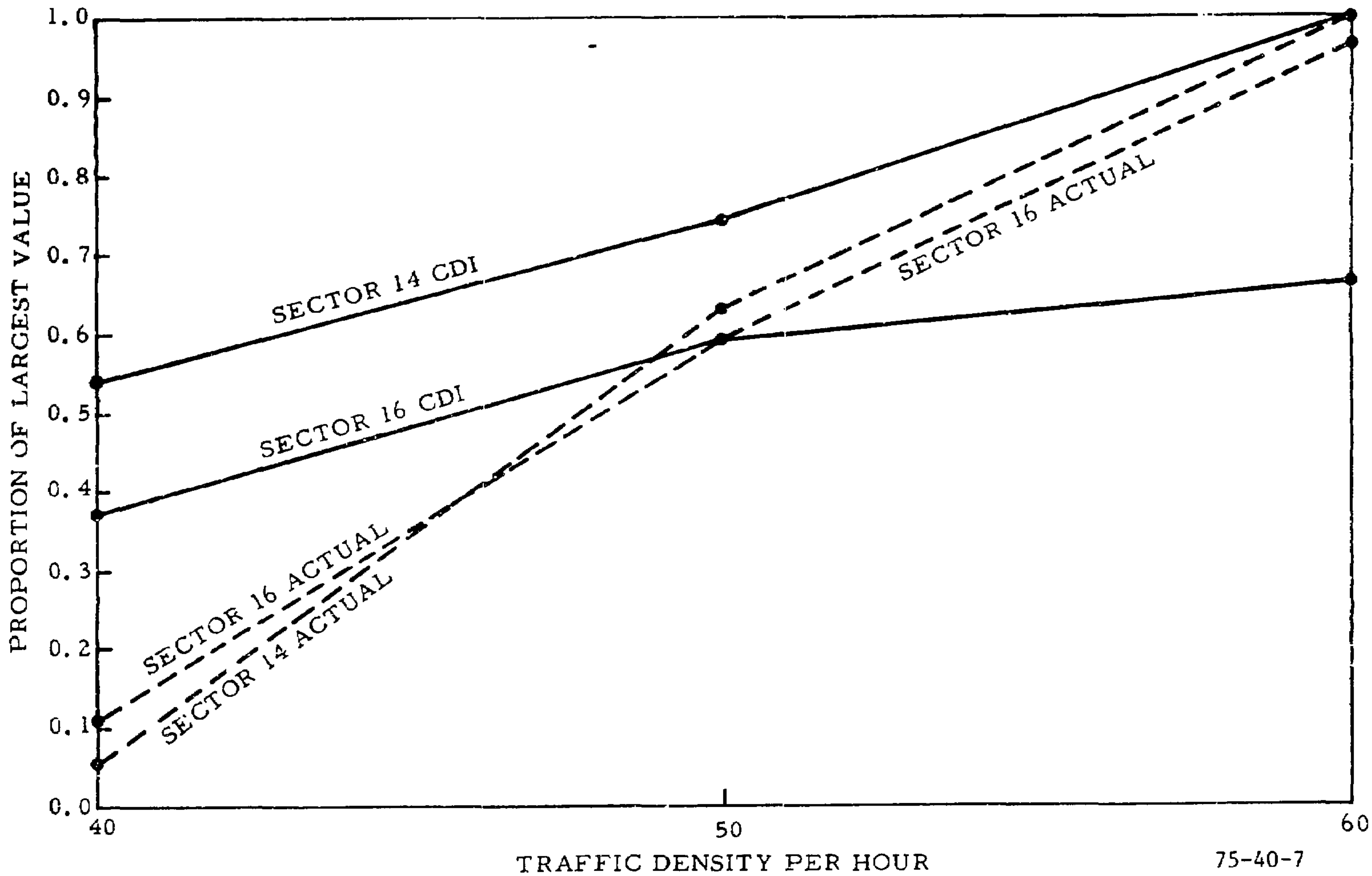
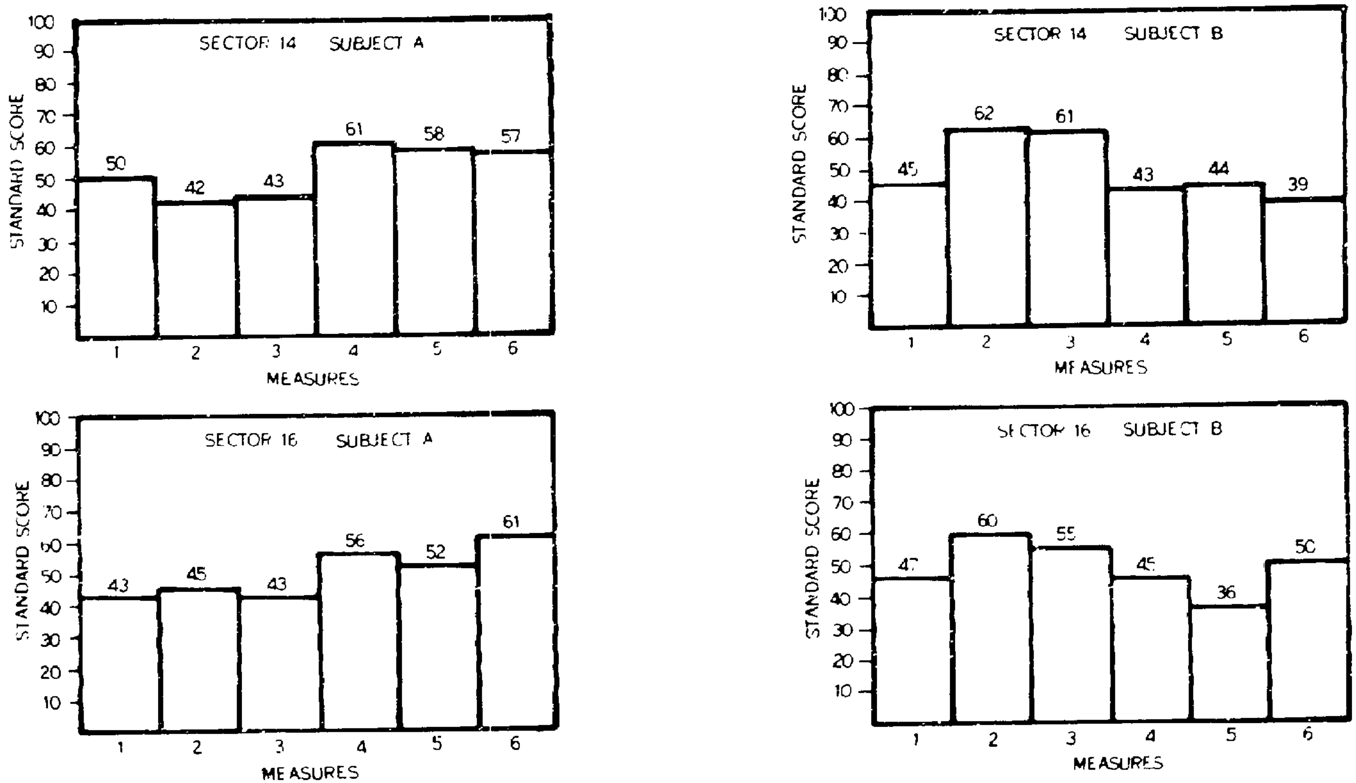


FIGURE 7. ACTUAL NUMBER OF DELAYS (AVERAGE) AS COMPARED TO STANFORD CONTROL DIFFICULTY INDEX (CDI)



MEASURES:

1 - NUMBER OF CONFLICTIONS/NUMBER OF AIRCRAFT HANDLED (1)
2 - NUMBER OF DELAYS/NUMBER OF AIRCRAFT IN SAMPLE (3)
3 - CUMULATIVE DELAY TIME/NUMBER OF AIRCRAFT IN SAMPLE (4)
4 - NUMBER OF COMPLETED FLIGHTS/NUMBER OF COMPLETABLE FLIGHTS (5)
5 - NUMBER OF AIRCRAFT HANDLED/NUMBER OF AIRCRAFT IN SAMPLE (8)
6 - CORRELATION HANDLED/DELAY TRANSFORM (9)

NOTE:
NUMBERS IN PARENTHESES REFER TO IDENTIFIER NUMBERS IN BASIC LIST OF MEASURE DEFINITIONS ON PAGES 8 AND 9.

75-40-8

FIGURE 8. STANDARD SCORE PROFILES OF TWO CONTROLLERS ON TWO SECTORS

SUMMARY OF RESULTS AND PERSPECTIVE

This experiment has contributed much information to guide future steps in the development of the Controller Performance Measurement (CPM) system. It has also reinforced old information. Reaffirmed, for example, is the perennially forgotten, or ignored, fact that there are wide differences among air traffic controllers in their ability to handle the identical traffic in the identical sector. Also demonstrated has been the fact that it is possible to measure the results of these differences in traffic-handling performance in a completely objective manner with only the computer doing the data collection.

The main contribution of this particular experiment appears to be the providing of an initial indication that sectors and their structure (three-dimensional) do not, if traffic density is controlled (i.e., kept constant or comparable), appear to be a very large contributor to control difficulty. They are factors to be considered, of course, but these are not major factors, compared to traffic density level. Perhaps the reason why this has not been realized is that it is difficult to think of a sector without its customary level of traffic.

On the other hand, it is necessary to forcefully point out that this PROBE experiment is only that; it gives an indication. The sample of subjects was limited and small, and the data points were few. The plan is that there will be opportunity to verify these conclusions on a broader base later during the process of developing and refining CPM.

There is a considerable amount of work yet to be done in developing a CPM system. Some redesign of measures would appear to be needed. Future experiments must more directly examine the problem of minimal optimal traffic sample length; 1 hour is certainly not enough. Even though these are probing experiments, not intended to be conclusive, but rather to guide future processes, more subjects should be obtained, if possible. Effort measurement in CPM (such as heart rate) and the meaning of differences in effort, as distinct from differences in performance, must be determined.

The next experiment planned in this series of small probe experiments will deal with the process of learning a given sector/density combination. Learning curves will be plotted for six consecutive sessions.

REFERENCES

1. Buckley, E. P. et al., A Comparative Analysis of Individual and System Performance Indices for the Air Traffic Control System, NAFEC, Rept NA-69-40, Atlantic City, 1969.

2. Henry, J. H. et al., Training of US Air Traffic Controllers, FAA DOT; Institute for Defense Analyses, Washington, D.C., IDA Report R-206, January, 1975.

3. Davis, Frederick B. et al., Standards for Educational and Psychological Tests, Washington, D. C., American Psychological Association, 1974.

4. Kirk, R. E., Experimental Design Procedures for the Behavioral Sciences, Brooks/Cole Publishing Co., Belmont, California, 1968.

5. Winer, B. J., Statistical Principles in Experimental Design, N. Y., McGraw Hill, 1962.

6. Hollander, M. and Wolfe, D. A., Nonparametric Statistical Methods, Wiley, New York, 1973.

7. Petracek, S. J. and Schmidt, D. K., Guidelines for the Evaluation and Redesign of ATC Enroute Sectors, Stanford Research Institute, Menlo Park, California. Technical Note TN-8181-74-1, Contract DOT-FAA 70WA-214, 1974.

8. Ratner, R. S. et al., The Air Traffic Controller's Contribution to ATC System Capacity in Manual and Automated Environments, Second Interim Report, Stanford Research Institute, Menlo Park, California, June 1972.

9. McNemar, Q., Psychological Statistics, Wiley, New York, 1962.