WADC TECHNICAL REPORT 59-120
ASTIA DOCUMENT NO. AD 213592

RESEARCH TO INVESTIGATE FACTORS AFFECTING MULTIPLE-TASK PSYCHOMOTOR PERFORMANCE

OSCAR S. ADAMS
RAPHAEL B. LEVINE
LOCKHEED AIRCRAFT CORPORATION

W. DEAN CHILES
AERO MEDICAL LABORATORY

MARCH 1959

CONTRACT NO. AF 33(616)-6050

1

WRIGHT AIR DEVELOPMENT CENTER
AIR RESEARCH AND DEVELOPMENT COMMAND
UNITED STATES AIR FORCE
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

NOTICES

- - - - -

Qualified requesters may obtain copies of this report from the Armed Services Technical Information Agency, (ASTIA), Arlington Hall Station, Arlington 12, Virginia.

- - - - -

This report has been released to the Office of Technical Services, U. S. Department of Commerce, Washington 25, D. C., for sale to the general public.

- - - - -

AD-213592

WADC TECHNICAL REPORT 59-120
ASTIA DOCUMENT NO. AD 213592

RESEARCH TO INVESTIGATE FACTORS AFFECTING MULTIPLE-TASK PSYCHOMOTOR PERFORMANCE

OSCAR S. ADAMS
RAPHAEL B. LEVINE
LOCKHEED AIRCRAFT CORPORATION

W. DEAN CHILES
AERO MEDICAL LABORATORY

MARCH 1959

CONTRACT NO. AF 33(616)-6050
PROJECT NO. 7184
TASK NO. 71582

WRIGHT AIR DEVELOPMENT CENTER
AIR RESEARCH AND DEVELOPMENT COMMAND
UNITED STATES AIR FORCE
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

McGregor & Werner Midwest Corp., Dayton, O.
40-874 & 875-1100-6-26-59

FOREWORD

This report summarizes research on multiple-task performance conducted by the Human Factors Research Department of the Operations Research Division, Lockheed Aircraft Corporation, Georgia Division, under the general direction of Dr. Jack A. Kraft, Department Manager, and Mr. Robert D. Roche, Division Engineer.

The work was supported by the Engineering Psychology Branch, Aero Medical Laboratory, Wright Air Development Center, on Contract No. AF 33(616)-6050, under Project No. 7184 entitled "Human Performance in Space Flight" and Task No. 71582 entitled "Performance Capabilities Under Space Environments." Dr. W. Dean Chiles served as Project Scientist and consultant.

This report was prepared by Dr. Oscar S. Adams, Project Leader and Principal Investigator, Dr. Raphael B. Levine, Assistant Project Leader, and Dr. W. Dean Chiles. The editorial, illustrative and clerical services were provided by Mrs. Joan Sikes, Technical Editor, Mrs. Joan Cree, Engineering Artist, Mrs. Betty Cole and Mrs. Marie Hutchins, Department Secretaries.

The check-out and maintenance of the experimental apparatus was assigned to the Electrical and Electronics Research Department, with Mr. Fred R. Willard serving as Department Manager. The work was under the supervision of Mr. Douglas B. Hatmaker, who was assisted by Mr. Arthur J. Pittock, Mr. William W. Hartsfield, and Mr. Sidney R. Smith.

The authors wish to express their appreciation to Mr. Richard E. Lincoln and Mr. Richard P. Smith who contributed significantly to the collection, reduction and analysis of the data. Also, we wish to acknowledge the assistance of Dr. M. Carr Payne, Jr., Department of Psychology, and Mr. Jack E. Wilson, Department of Electrical Engineering, Georgia Institute of Technology, who were employed as College Faculty Associates during the Summer, 1958, and the contribution of Dr. Hudson Jost, Professor of Psychology, University of Georgia who served as a consultant.

ABSTRACT

Fifteen male subjects participated in an experiment designed to study: (a) the 24-hour test-retest reliability and intercorrelation of a battery of seven performance tasks; (b) the effects of performing simultaneously various combinations of physically compatible tasks; (c) the 24-hour test-retest reliability and intercorrelation of a battery of four psychophysiological variables; and (d) the relation of psychophysiological criterion measures to performance criteria and to conditions of task presentation.

The results indicate moderately high reliability and sensitivity for five of the performance task criteria, but a negligible amount of intercorrelation. While no significant correlation was obtained between performance scores and measures of psychophysiological activity, several of the psychophysiological measures showed significant reliability, intercorrelation, and sensitivity to conditions of rest and performance activity.

PUBLICATION REVIEW

This report has been reviewed and is approved.

FOR THE COMMANDER:

Walter F. Grether

WALTER F. GRETHER
Director of Operations
Aero Medical Laboratory

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

INTRODUCTION

The past twenty years have witnessed the development of increasingly complex weapon systems. This trend has produced two effects. First, a considerable amount of technological information is now available for the solution of many complicated engineering problems which have hitherto precluded travel in space. Thus, there are (or soon will be) in the Air Force inventory, weapon systems capable of performance in regions beyond the earth's atmosphere. Second, this trend has created a primary need for quantitative information concerning the limitations, capabilities and reliability of the human component because of the fact that one of the major unknowns in advanced weapon-system performance will be the human operator. More specifically, designers need to know precisely what human functions can be utilized in the operation of space vehicles or other advanced man-machine systems. Hence, it will be necessary to determine the reliability with which such functions can be performed, as well as the level of performance that can be maintained over long periods of time. Of equal importance is the need for information concerning the simultaneous performance of tasks which involve different psychological functions. Although, when taken individually, certain tasks may be performed reliably at an acceptable level, our present knowledge does not permit us to predict the level of performance that can be achieved and maintained when the operator is required to perform several tasks simultaneously.

The number of previous studies which are applicable to this problem is small, since most investigations which might seem relevant have dealt primarily with single (usually simple) tasks. For example, in establishing selection batteries, the usual procedure is to choose that combination of tests which yields the highest multiple correlation with a predetermined criterion measure when the component tests in the battery are administered individually.

Studies concerning the learning of skills are important to the present investigation, but the major interest here is with the maintenance rather than the acquisition of skills, as well as with the mutual interference of simultaneous activities. A further distinction can be made between our interest and that of many others; for example, investigations of task load have shown that, as the number and/or sources of signals are increased, there is a resulting decrease in operator efficiency. In general, however, such studies (8) have imposed a task load by increasing the number of signals or signal sources within a specific skill dimension, whereas in the present framework the number of skill dimensions is of interest.

Bartlett and his colleagues (5) and Hauty and Payne (16) have studied the effects of fatigue-inducing conditions on simultaneously performed tasks. Multiple-display problems were presented by the former in the Cambridge Cockpit (10) and by the latter in the USAF SAM Multi-dimensional Pursuit Test (20). The reported results, some of which are conflicting, will not be discussed here except to say that there is some evidence that general fatigue decrement may be influenced by intertask factors. It is possible that some of the differences in results can be attributed to a difference in the number of skill dimensions

---

Manuscript released by the authors 20 March 1959, for publication as WADC Technical Report.

involved, as well as to the number and arrangement of task displays within a given dimension.

A study by Baker, Wylie, and Gagne (3) is somewhat applicable to this problem, although it is primarily concerned with skill acquisition. They have shown that performance time increases sharply when an interfering task (memory for verbal instructions) is introduced at any one of six stages during the learning of the SAM Complex Coordination Test. Following the introduction of interference, performance time for the interference groups was found to increase from 20 to 40 percent over that for a control group which did not experience interference until the last stage. Further, the effects of interference were never overcome within the range of practice used.

The psychophysiological approach to problems of performance, while thus far lacking a detailed theoretical justification, is based on a good deal of empirical fact and is being more and more widely used. Many of the symptoms of tension are reflected in electrical and other physiological changes which are amenable to convenient recording techniques and at least semi-quantitative mathematical analyses. Questions concerning the desirability of using physiological measurements, the advantages of taking simultaneous multiple measurements and the selection of specific variables measured were discussed in detail in a previous report (1). The answers to these questions can be summarized here as follows:

A number of studies have shown that various physiological measures provide for the assessment of emotions, tensions, and fatigue in the human (2, 4, 6, 7, 9, 12, 18, 19, 21, 23, 24). Since factors such as task complexity, confinement, and performance duration can be expected to influence these subjective variables, it is desirable that whatever reasonable techniques are available for their measurement should be employed. In view of Lacey's work (18, 19) on intersubject differences for various psychophysiological measures (the fact that different individuals show variations in different bio-electric continua) it is necessary that several mechanisms be assessed simultaneously. Although the question as to which specific measures should be used is too bulky for consideration here, its answer revolves around factors such as amount of background data and ease of interpretation.

The purpose of the present investigation was fourfold: First, to provide reliability data on a test battery consisting of four different monitoring tasks, an arithmetic computation task, a tracking task, and a pattern perception task. Second, to provide reliability data on a group of psychophysiological measures involving heart rate, breathing rate, skin temperature and skin resistance. Third, to provide measures of possible interrelationships between the psychophysiological and performance measures. And fourth, to provide data on the simultaneous performance of various combinations of these tasks.

The specific performance tests selected for investigation were chosen because of their relationships to tasks which, it is presumed, will be required of the human operator in future weapon systems.

METHOD

TEST EQUIPMENT

Subjects were tested in a five-man, advanced flight station mock-up which has been described in a previously published report (1). In order to minimize visual distraction and social interaction among subjects, curtains were placed around each of the subject stations. A random noise signal with an intensity level of 85 decibels was maintained during the test periods to mask outside auditory stimuli. One-way vision screens and closed-circuit television provided direct observation of each subject at all times.

SUBJECTS

The subject sample consisted of 15 college students enrolled in the Lockheed Engineering Cooperative Training program who had volunteered to participate in the study. They were paid as a part of other regular work. The subjects ranged in age from 18.4 years to 23.0 years. On the average, they had completed 3.7 quarters of college work.

PERFORMANCE TASKS

A slight modification of the battery of performance tasks described previously (1) was used throughout the present study. The battery consists of seven tasks which were selected on the basis of three criteria. The criteria established were that the tasks should be (a) representative of the psychological factors important to a wide variety of operator performance duties, (b) minimally influenced by specific past experience, and (c) show promise of being sensitive to the experimental conditions to be imposed.

The seven tasks finally selected represent psychological functions such as perceptual-motor coordination, mental computations, pattern discrimination, monitoring and vigilance. The performance tasks are built into an 11-inch by 28-inch panel (see Figure 1). This panel is mounted with the surface tilted away from the subject at an angle of 12 degrees from the vertical so that the average subject views the displays at a distance of approximately 20 inches, the top of the panel being about shoulder high (see Figure 2).

Each of the specific tasks is described in the following paragraphs.

Compensatory Tracking

The tracking display consists of a standard cross-pointer indicator which is mounted in the upper center of the instrument panel. Movements of the horizontal and vertical pointers are programmed by separate signal outputs from two pairs of low frequency noise generators. The subject's task is to keep the pointers centered on vertical and horizontal reference marks at the edge of the scale by operating a formation-type stick control located between his legs. A continuous measure of performance is obtained for each pointer by recording the error on magnetic tape as the difference between the appropriate input and output voltages. The recorded signal is subsequently fed to

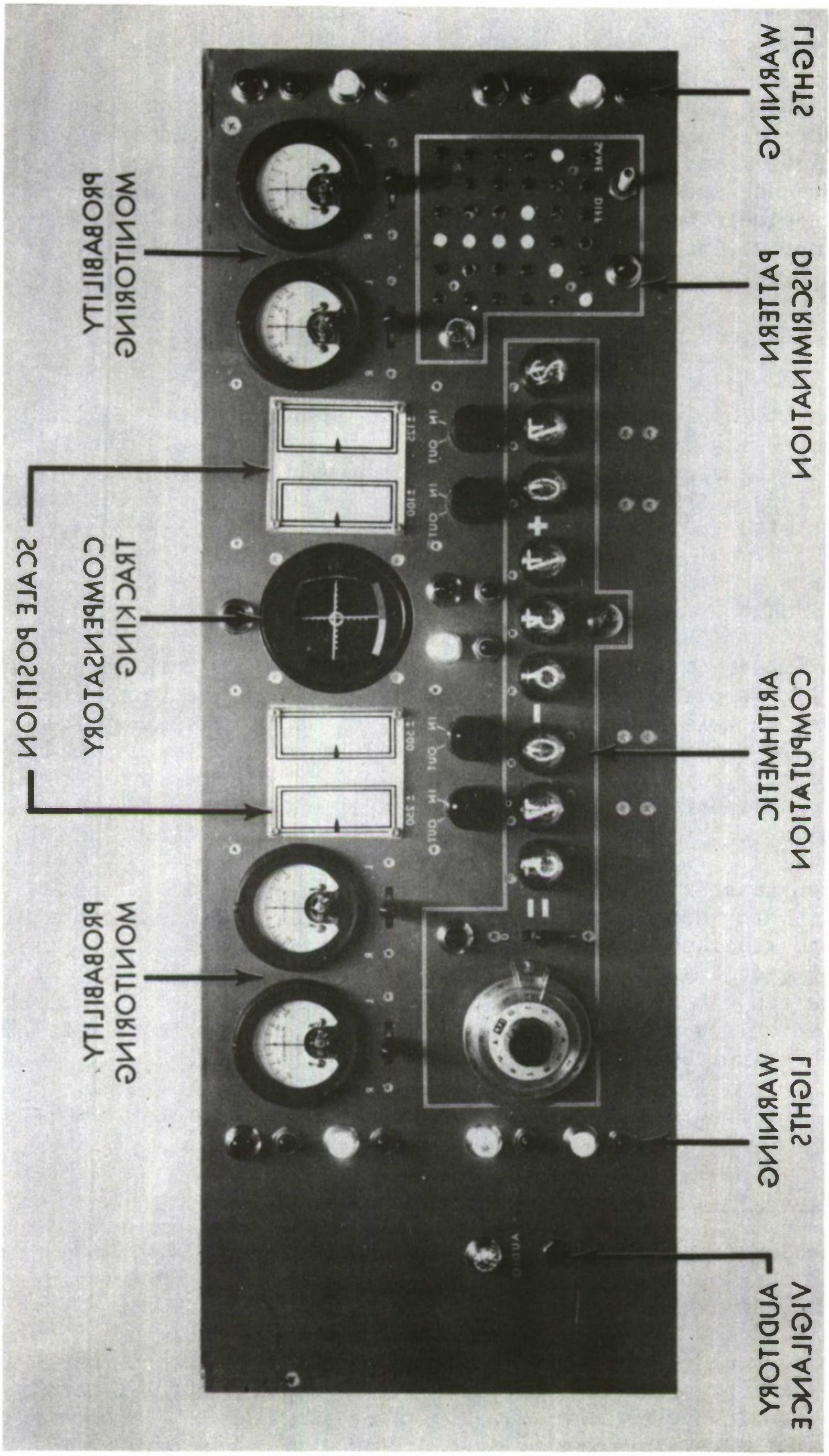


Figure 1. Subject's Performance Panel

Figure 2. View of Subject's Work Station and Performance Panel Located on Starboard Side of Compartment. Curtain has been removed.

an EASE Analog Computer which calculates the root mean square (RMS) error for each of the two dimensions for each five-minute interval of performance.

Arithmetic Computation

The display for this task consists of nine single-digit numerical indicator tubes, grouped by threes and arranged horizontally along the lower central portion of the instrument panel. This arrangement provides for the presentation of three 3-digit numbers, the first two numbers being separated by a plus sign painted on the panel and the second and third by a minus sign.

The subject is required to subtract the third three-digit number from the sum of the first two. He records the units, tens, and hundreds places of his answer on three concentric knobs located at the extreme right on the lower portion of the panel and then indicates whether the thousands place is a 1 or a 0 by activating a three-position, spring-centered, toggle switch. Movement of the switch automatically records the response as being right or wrong and turns on a blue light on the subject's panel if the answer is correct.

A basic program of 100 problems is presented at a constant rate of three per minute, the number of problems each subject solves correctly being recorded each 15 minutes.

Pattern Discrimination

A 6 x 6 matrix of lights, mounted in a four-inch square array at the lower left corner of the instrument panel is used to display the patterns. This matrix of 36 lights is wired so that any number of lights from 1 to 36 can be presented simultaneously in any combination.

A punched tape programmer provides 120 different pairs of patterns which are presented at a constant rate of two pairs per minute. The first member of the pair is presented for five seconds, followed by a five-second off period. The second member of the pair is then presented for five seconds, followed by a 15-second period in which the subject indicates whether his judgment is "same" or "different," by throwing a three-position, spring-centered, toggle switch in the appropriate direction.

In order to force the utilization of figure or shape in making judgments, the second member of the pair is rotated 90°, 180°, or 270° from the orientation of the first member. In one-half of the presentations some aspect of the configuration itself is changed in the second member of the pair.

The illumination of a blue light indicates to the subject that his answer is correct. The criterion measure is the number of correct responses for each 15-minute period of performance.

Auditory Vigilance

"Beeps" of a 1175-cps tone are presented once every 0.56 seconds through a single earphone worn by the subject. The tone is normally on for 0.12 seconds, and off for 0.44 seconds. The critical signal to be detected is a single off-period of 0.55 seconds which occurs an average of four times per hour. When the subject thinks that he has detected this change, he depresses a pushbutton which is located at the edge of his instrument panel. The subject is not provided with feedback of knowledge of results, as is the case in the other tasks.

The number of correct responses, as well as the total number of times the subject has depressed his pushbutton are recorded. In order for a response to be scored as correct, the subject must respond within 30 seconds after the occurrence of a critical signal.

Scale Position Monitoring

This display consists of four vertically oriented moving-tape scales located in the upper portion of the panel, two on either side of the cross-pointer indicator. The scales are printed on transparent 35-mm film which is back-lighted to project onto a translucent plate. Each of the scales contains a different number of scale divisions representing different units, and each scale must be monitored with respect to different tolerances.

Movement of the scales is controlled by four separate programmers, each containing a program of 13.5-minutes' duration. At the end of this cycle the programs are re-designated with respect to which will control a given scale. Unknown to the subject, only one scale is permitted to be out-of-tolerance at a time; the average total out-of-tolerance time is about four minutes per program for all scales combined for a given subject.

A two-position switch is located below each scale. The subject signals out-of-tolerance indications by turning the knob to the OUT position and signals the return of a scale to tolerance by turning the knob to IN. Two separate time scores are recorded at each cycle of the program. The first of these is the total time for which any one of the scales was out of tolerance without being detected; the second is the return time, i.e., the total time for which the subject indicates that there is an out-of-tolerance reading on any of these displays when, in fact, none exists.

Warning Lights Monitoring

The display for this task consists of five red and five green indicator lights which are located in pairs on the side of the panel and below the tracking display. The subject is required to turn any given green light on if it goes off, and any red light off if it comes on, by pressing the appropriate pushbutton located beneath the light in question.

An average of ten non-normal indications, five red and five green, are presented at random over the hour of performance. The accumulated response times to green light and to red light non-normal indications are recorded on separate 0.1-second timers, one for all red lights and one for all green lights.

Probability Monitoring

This task, which is a modification of one used by Howland (17), consists of four displays with semicircular scales. A pointer on each display is driven by a low-frequency noise generator, the pointer settings being normally distributed with a mean of zero (12 o'clock position) and a standard deviation of 25 scale units. Every five minutes a programmer shifts the mean (but does not change the variance) of the distribution 25 units to the right or left by introducing an appropriate bias to one of the four dials selected at random. The rate at which the introduction of the bias moves a given pointer to its new location is essentially indistinguishable from the usual rates of pointer movement. Therefore, the subject is required to integrate the positions of a given pointer over time in order to decide whether or not a bias is present. When the subject suspects that a particular pointer is biased in a given direction, he moves the three-position, spring-centered, toggle switch associated with that dial in that direction, whereupon the pointer moves to the mean of its distribution, i.e., a reading of -25, 0 or +25. Thus the subject gets immediate feedback with respect to the appropriateness of his decision. If a bias is present, release of the switch resets the system to a zero-bias condition and the program continues.

On the average, nine biases are introduced in the course of an hour. At the end of each five minutes a record is made of whether a bias which may have been present was detected, of the number of false responses, and of the time required to detect a bias if introduced.

PSYCHOPHYSIOLOGICAL MEASURES

Each of the four psychophysiological measures finally selected for use in this study is described briefly below.

Skin Resistance

Skin resistance is measured by the Fels Dermohmmeter which utilizes a direct current of 70 microamperes through a palmar zinc-to-zinc-sulfate-in-agar electrode. The return is through a large lead electrode shaped to the homolateral forearm which is prepared by rubbing it with electrode paste. This instrument is essentially a self-balancing Wheatstone bridge, which presents its readings in three parts. The first part gives the position of a range switch calibrated in hundreds of thousands of ohms; the second part gives the position of another range switch calibrated in tens of thousands of ohms; the third part is a meter indication with a range of plus and minus 6000 ohms. For recording purposes, the instrument delivers information concerning the meter reading and the position of the range switches.

Two measures are chosen to characterize each five-minute interval of record. The first of these is a skin resistance "level" measure; it is the mean of 20 instantaneous values of resistance, one for each 15 second sub-interval on the record. The second, which takes any value from 0 through 20, is the number of these above mentioned 15-second sub-intervals during which at least one galvanic skin response occurred, i.e., a decrease in resistance of 1,000 ohms or more which occurs within a three-second period.

Skin Temperature

The Yellow Springs Instrument Company Telethermometer is used to measure skin temperature. The sensing element of this instrument is a thermistor embedded in a 3/8-inch diameter button held against the skin by a piece of tape. The telethermometer is essentially a D.C. Wheatstone bridge which automatically balances itself in steps against the resistance of the thermistor. The residual imbalance is amplified and drives a meter having a full-scale deflection which covers a range of 1.2 degrees Centigrade. When the self-balancing circuit is activated it drives a range switch up or down in one-degree steps. For recording purposes, the instrument delivers outputs which reflect the meter reading and the range switch position.

Two measures are used to characterize each five-minute interval recorded: the arithmetic mean value of the temperature during the five minutes (measured each 30 seconds), and the variance of the changes from one 30-second reading to the next. The former score is taken to characterize the level of skin temperature; the latter is a measure of the fluctuation in the skin temperature.

Heart Rate

This instrument measures the duration of each individual cardiac cycle (R-R interval) and presents its reciprocal as a meter reading. An output voltage proportional to this reading provides for an oscillographic record of cardiac rates to the nearest beat per minute.

Two measures are taken to characterize each five-minute interval of record, namely, heart rate level and heart rate fluctuation. The five-minute interval is divided into five-second subintervals, and the maximum and minimum rates observed during each of these subintervals is noted. From these values, heart rate level is computed as the mean of the 60 maximum plus the 60 minimum readings for each five-minute interval. The fluctuation measure is the number of subintervals in which the maximum rate exceeds the maximum rate of the immediately preceding subinterval by nine beats per minute or more. Although the fluctuation measure can theoretically take values from 0 through 59, in practice the rate does not continue to increase by nine beats per five seconds for more than perhaps 20 seconds at a maximum.

Respiration Rate

A respirometer designed in this laboratory is used for the measurement of the breathing movements. It utilizes information received from a pair of carbon buttons which are encased in a potting compound and placed in an elastic belt that holds them against the subject's body. The buttons are alternately compressed and released by the respiratory motions of the abdomen or chest, and the resulting variations in resistance drive a meter, the movement of which follows the respiratory motions.

Two measures are used to characterize each five-minute interval of record: respiration rate level and respiration rate fluctuation. The level measure is the average number of breaths per 15 seconds for each five-minute interval. The fluctuation measure is the number of 15-second subintervals in which the number of breaths differs by ±2 or more from the count in the immediately preceding subinterval.

PSYCHOPHYSIOLOGICAL RECORDING

The sensing elements for the psychophysiological measures are applied as a harness attached to an elastic belt fastened about the subject's waist. A sketch of the pickup arrangement is shown in Figure 3.

The electrical leads from all the pickups are joined to form a cable which terminates in a multi-pin plug which in turn is plugged into a receptacle in the seat of the experimental capsule. Pockets in the elastic belt contain the carbon buttons for the respiration pickup, and separate leads from the cable go to the skin temperature thermistor affixed to the forehead, to the skin resistance electrodes on the left arm, and to the cardiotachometer electrodes on the chest. Locations for the cardiotachometer electrodes are chosen for each individual subject so as to pick up the largest possible voltage and, thus, maximize the reliability of the recordings. The electrodes are held in place by an auxiliary rubber strap, the contact being improved by the use of electrocardiographic paste.

The psychophysiological data are recorded on 30 galvanometers of a Consolidated Electrodynamics Corporation multichannel oscillograph, six channels for each of the five subjects in a given group. The tremendous amount of data recorded on these measures (over 830,000 data points) dictated that some sort of sampling procedure be used in analyzing the psychophysiological variables. The procedure finally settled on was to select for analysis a five-minute interval during the pre- and inter-test rest periods for the control data and, for the performance periods, a five-minute interval at the beginning (the 1st through the 5th minutes), one at the middle (25th through 30th), and one at the end (55th through 60th). Appropriate shifts in the selection were employed if either subject-movement or apparatus-induced artifacts were apparent.

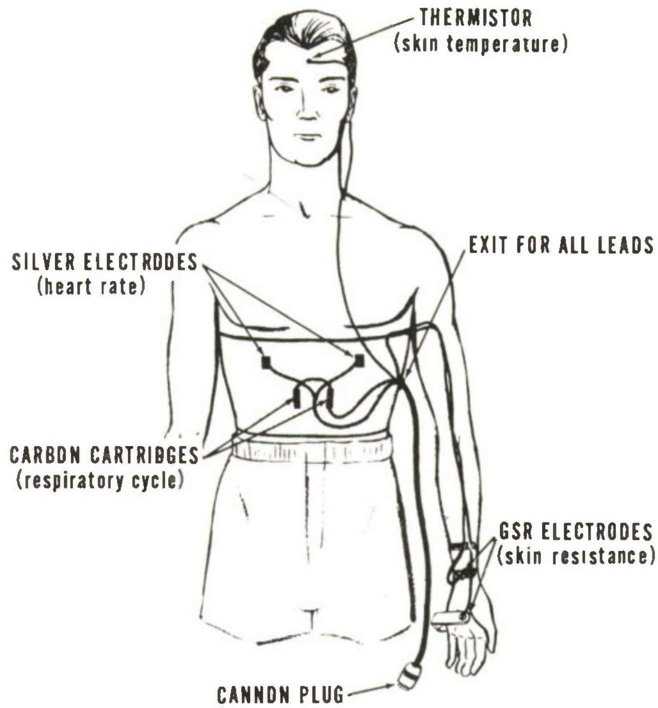


Figure 3. Physiological Harness

PROCEDURE

The 15 subjects were divided into three groups of five subjects each. Within a group, all five subjects were presented the same tasks and tested simultaneously. The test program was divided into two Phases (I and II), with each group participating first in Phase I and next in Phase II. Observations were continued over a period of seven days, the first five days being allocated to Phase I and the last two to Phase II. For all groups the test program commenced on a Thursday and ended on the Friday of the following week. This permitted the last three days of Phase I and the two days of Phase II to be consecutive.

Phase I: Learning and Test Reliability

During Phase I, the subjects were familiarized with each performance task and were given three hours of practice on each task spaced over the first three days to allow performance to approximate an asymptote.

The last two days of Phase I were the same as the first three days except that, in addition, the psychophysiological measures were recorded prior to and during performance. The performance and psychophysiological scores obtained during these two days were used to compute test-retest coefficients of reliability. Each task was presented for one hour each day, making a total of five hours of performance testing in any one day. Order of task presentation was counterbalanced within the limits set by the number of test groups (see Table 6, Appendix A). The tracking, arithmetic, and pattern discrimination tasks were presented individually; the auditory vigilance and scale position tasks were always presented together as were the warning lights and probability monitoring tasks. Thus, for experimental design purposes, in effect only five tasks were used.

Successive testing intervals were separated by rest periods at least 30 minutes long, during which time the subjects were allowed to relax. On Days 4 and 5 electrode paste was replenished during these rest periods, and a five-minute psychophysiological control record was taken.

Phase II: Multiple-Task Presentation

It was assumed that the activities required of the subject by the tracking, arithmetic and pattern discrimination tasks were mutually incompatible at least for the purposes of this study. Imposing this restriction reduces the number of test combinations to the following 7 two-test and 3 three-test arrangements: (a) tracking, auditory vigilance and scale position (T, AV & SP); (b) tracking, warning lights and probability monitoring (T, WL & PM); (c) arithmetic, auditory vigilance and scale position (A, AV & SP); (d) arithmetic, warning lights and probability monitoring (A, WL & PM); (e) pattern discrimination, auditory vigilance and scale position (P, AV & SP); (f) pattern discrimination, warning lights and probability monitoring (P, WL & PM); (g) auditory vigilance, scale position, warning lights and probability monitoring (AV, SP, WL & PM); (h) tracking, auditory vigilance, scale position, warning lights and probability monitoring (T, AV, SP, WL & PM); (i) arithmetic, auditory vigilance, scale position, warning lights and probability monitoring (A, AV, SP, WL & PM); and (j) pattern discrimination, auditory vigilance, scale position, warning lights and probability monitoring (P, AV, SP, WL & PM). These combinations were presented the last two days of testing for a given group, i.e., five combinations on Day 6 and five on Day 7. As before, a 30-minute rest period was given between combinations, and counterbalancing, within the limits of the sample size, was used (Table 7, Appendix A). The psychophysiological measures were recorded with the same procedure used on Days 4 and 5.

RESULTS

RELIABILITY OF MEASURES

Performance Tasks

For each criterion measure a product-moment coefficient of correlation was computed between the appropriate scores obtained on Day 4 and the associated scores obtained on Day 5. All correlations were based on an N of 15 except as noted for the tracking task scores. A summary of all of these reliability coefficients is presented in Table 1.

Reliability coefficients were computed for the tracking task using 10 subjects; the five subjects in Group I were dropped after it had been determined that equipment malfunction had been present during that portion of the testing. These coefficients showed the day-to-day variability to be considerable on all three measures, the r's for horizontal, vertical and total RMS error being .156, -.152 and .202 respectively. However, when the odd and even 5-minute intervals were correlated within days, reliability coefficients of .942 for Day 4 and .878 for Day 5 (both of which are significant) were obtained.

TABLE 1

Reliability Coefficients for Performance Measures

Criterion Measure	r	Criterion Measure	r
Tracking		Scale Position	
Horizontal Error	.156	Detection Time	.758**
Vertical Error	-.152	Return Time	-.105
Total Error	.202		
		Warning Lights	
Arithmetic		Response Time (Red)	.567*
Number Correct	.787**	Response Time (Green)	.472
Pattern Discrimination		Probability Monitoring	
Number Correct	.740**	Percent Correct	.781**
		Detection Time	.807**
Auditory Vigilance		False Responses	.926**
$\frac{\text{Number Correct}}{\text{No. Signals + No. False}}$	.094	* P ≦ .05	
		** P ≦ .01	
No. False + No. Missed	.158		

The reliability coefficient obtained for the arithmetic task was .787 which permits the rejection of the hypothesis of zero correlation at the 1% level.

The correlation found between Days 4 and 5 for the pattern discrimination task was .740 which is significant at the 1% level.

There was apparently no relationship between the performance of the auditory vigilance task on Day 4 and that on Day 5, irrespective of the measure used.

The detection-time measure for the scale position task was quite reliable showing a significant reliability coefficient of .758. The return time measure appears to represent no stable relationship over time.

The reliabilities of the warning light task were shown to be marginal, the coefficients being .472 (not significant) for the response time to the green lights and .567 (5%) to the red lights.

The coefficients computed for the three measures for the probability monitoring task were .781 for accuracy of response, .926 for false response and .807 for detection time, all of which are significant at better than the 1% level of confidence.

Psychophysiological Measures

Reliability coefficients (product moment correlation) were computed using the average of the five resting intervals for Days 4 against those for Day 5; a second set of coefficients was computed using the analogous data for Days 6 and 7 (Table 2). Upon inspection of this table, it can be seen that for Days 4

TABLE 2

Reliability Coefficients for Psychophysiological Measures During Resting Periods

Measure	Day 4 vs. 5		Day 6 vs. 7	
	Level	Fluctuation	Level	Fluctuation
Skin Resistance	.511*	.861**	.631**	.804**
Skin Temperature	.639**	-.020	.757**	.220
Heart Rate	.843**	.323	.272	.067
Respiration Rate	.911**	.426	.901**	.029

* P = .05
** P ≤ .01

and 5, all of the reliability coefficients for the level measures are significant--skin temperature, heart rate, and respiration at the 1% level of confidence and skin resistance at the 5% level. The level measures for Days 6 and 7 show reliabilities of .757 (1%), .631 (5%), .272 (not significant) and .901 (1%) for skin temperature, skin resistance, heart rate and respiration respectively. For the fluctuation measures, only skin resistance showed a significant degree of reliability, and it was significant at the 1% level.

An additional set of reliability coefficients was computed for Days 4 and 5. These coefficients were based on the average of the three sample periods which were analyzed for the hour of performance on each task on Day 4 and Day 5. The correlations will not be discussed but are presented in Table 8, Appendix B, for the psychophysiologically oriented reader.

INTERCORRELATIONS

In order to obtain estimates of the extent to which the various performance task measures are interrelated, the correlation of each measure--both between and within tasks--with each other measure was computed. Of this matrix of 66 intercorrelation coefficients (Table 9, Appendix B) only 3 were significant. The first of these was that between the two auditory vigilance measures (r = -.964). This negative correlation is probably inherent in the method of calculating the two scores, i.e., the number of false responses appears in the numerator of one and the denominator of the other. The correlation of .750 between response time to red lights and that to green lights is certainly to be expected. The negative correlation, -.711 (1%), between response time and number of false responses for the probability monitoring task is surprisingly close to that (-.707) obtained by Howland (17) for the performance of 34 subjects on a very similar task.

Of the 28 intercorrelations computed on the eight psychophysiological measures, only those among the four fluctuation measures (Table 3) were significant. In the complete matrix (presented in Table 10, Appendix B) none of the correlations between pairs of level measures or between paired level and fluctuation measures was significant.

TABLE 3

Intercorrelations for Psychophysiological Fluctuation Measures

Measure	1	2	3	4
1. Skin Resistance		.544*	.588*	.774**
2. Skin Temperature			.621**	.773**
3. Heart Rate				.913**
4. Respiration Rate				

* P ≤ .05
** P ≤ .01

Of the 96 intercorrelations of performance and psychophysiological measures, three were significant at the 5% level and one at the 1% level. It will be remembered that, for this number of correlations, we would--on the basis of chance alone--expect five values to reach the 5% level and one of these five to reach the 1% level of confidence.

VARIATIONS IN TASK COMBINATIONS

Performance Tasks

The purpose of this phase of the study was to determine what differences, if any, would be observed between the level of performance achieved under conditions of single-task performance and that achieved under the several combinations involving multiple-task performance. To ascertain whether or not any of these differences were statistically significant, Friedman's analysis of variance test for ranked data (13) was applied to each of the criterion measures. The tracking, arithmetic, and pattern discrimination tasks each provided four scores to be ranked, one from Day 5 and three from Days 6 and 7. The score from Day 5 was for single-task performance, and the three scores from Days 6 and 7 were for the combinations involving the monitoring tasks (cf p. 11). In the case of each of the monitoring tasks (auditory vigilance, scale position monitoring, warning lights monitoring, and probability monitoring) there were eight scores to be ranked, one from Day 5 (single task) and seven from Days 6 and 7 (multiple task). The Chi-squares yielded by the Friedman test are presented in Table 4.

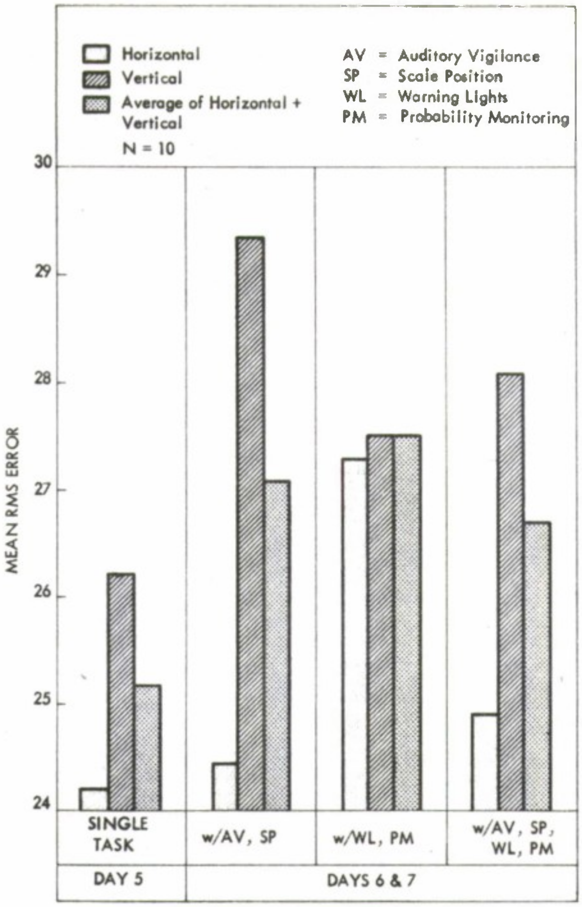
TABLE 4

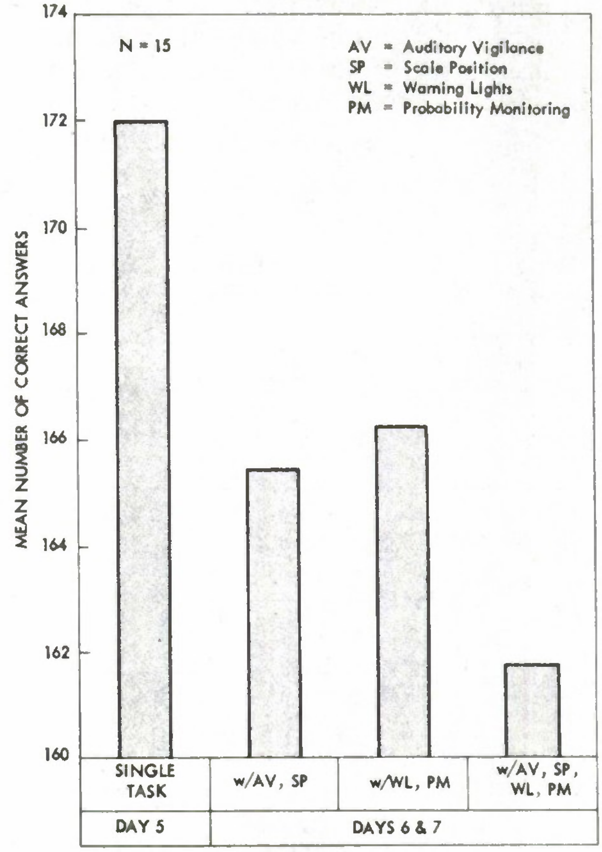
Results of Friedman's Test of Significance of Differences Among Task Conditions for Each Performance Criterion Measure

Criterion Measure	df	χ_r^2	P-Level
Tracking (Horizontal + Vertical Error)	3	2.04	--
Arithmetic (Number Correct)	3	12.66	<.005
Pattern Discrimination (Number Correct)	3	1.10	--
Auditory Vigilance (No. Correct/No. Signals + False)	7	12.14	<.10
Auditory Vigilance (No. False + No. Missed)	7	9.30	--
Scale Position Monitoring (Detection Time)	7	15.42	<.05
Scale Position Monitoring (Return Time)	7	20.34	.005
Warning Lights (Red-Light Response Time)	7	16.14	<.02
Warning Lights (Green-Light Response Time)	7	27.62	<.001
Probability Monitoring (No. Correct/No. Biases)	7	12.02	.10
Probability Monitoring (Detection Time)	7	16.86	<.02
Probability Monitoring (No. False Responses)	7	14.46	<.05

The mean RMS scores for horizontal and vertical tracking error are shown in Figure 4 together with the RMS of the average of both errors. The Chi-square for the various combinations of task performance failed to reach significance (see Table 4).

The mean numbers of arithmetic problems solved correctly during single-task performance on Day 5 and during multiple-task performance on Days 6 and 7 are shown in Figure 5. This plot indicates that the performance levels under conditions of multiple-task performance were lower than that for single-task performance. The Friedman test showed that these differences are significant at the .005 level.





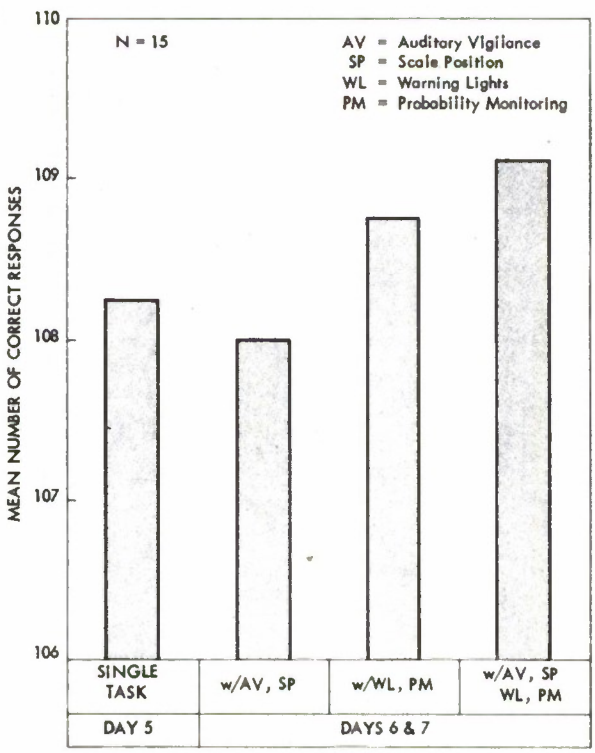


Figure 4 (ABOVE LEFT). Comparison of Tracking Scores Obtained on Day 5 and During Performance with Other Tasks, As Indicated, on Days 6 and 7.

Figure 5 (ABOVE RIGHT). Comparison of Arithmetic Scores Obtained on Day 5 and During Performance with Other Tasks, As Indicated, on Days 6 and 7.

Figure 6 (LEFT). Comparison of Pattern Discrimination Scores Obtained on Day 5 and During Performance with Other Tasks, As Indicated, on Days 6 and 7.

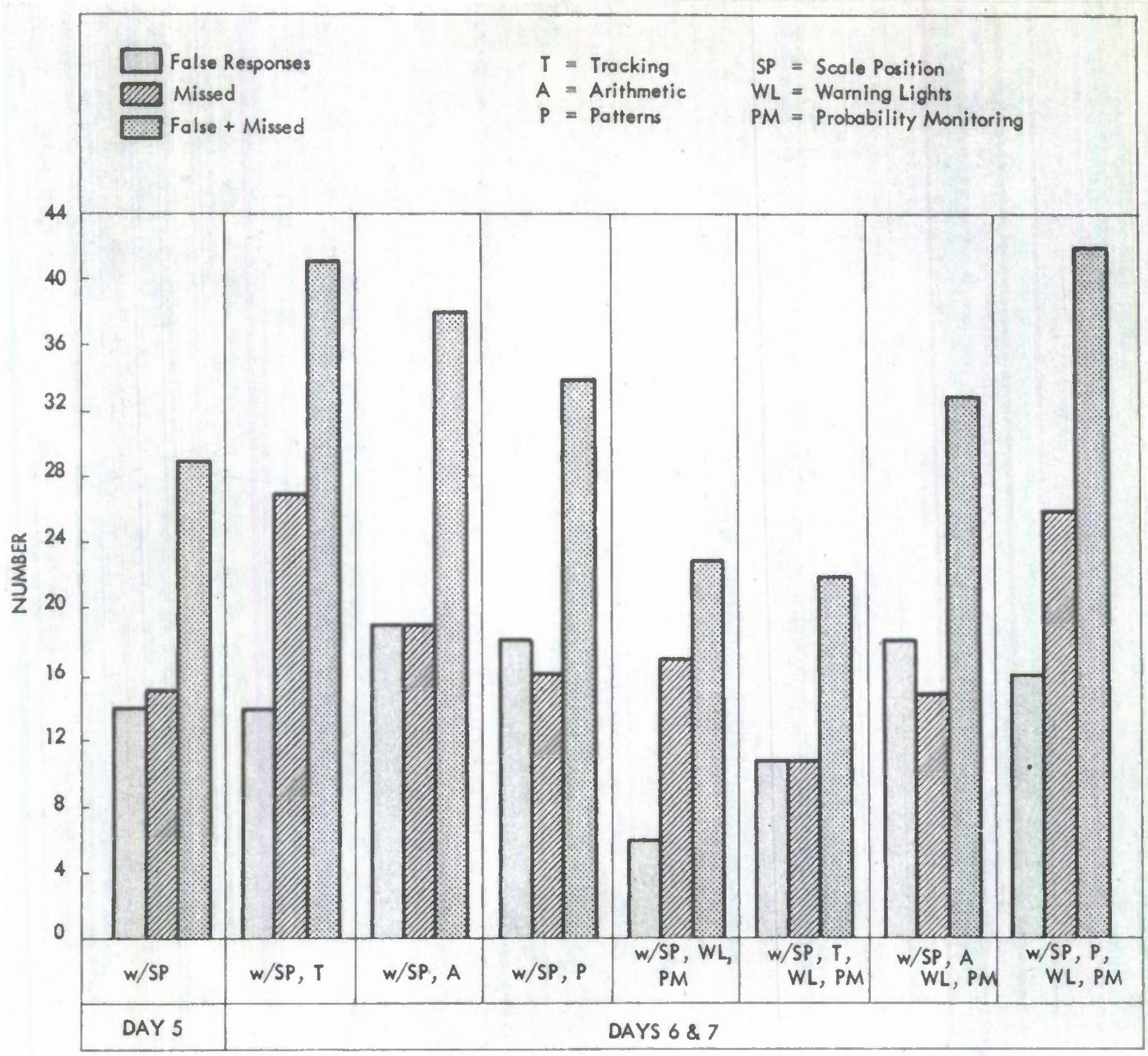
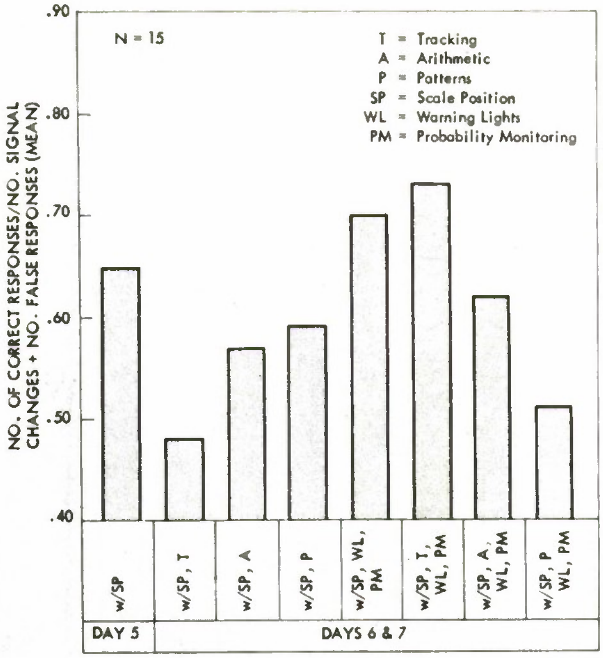
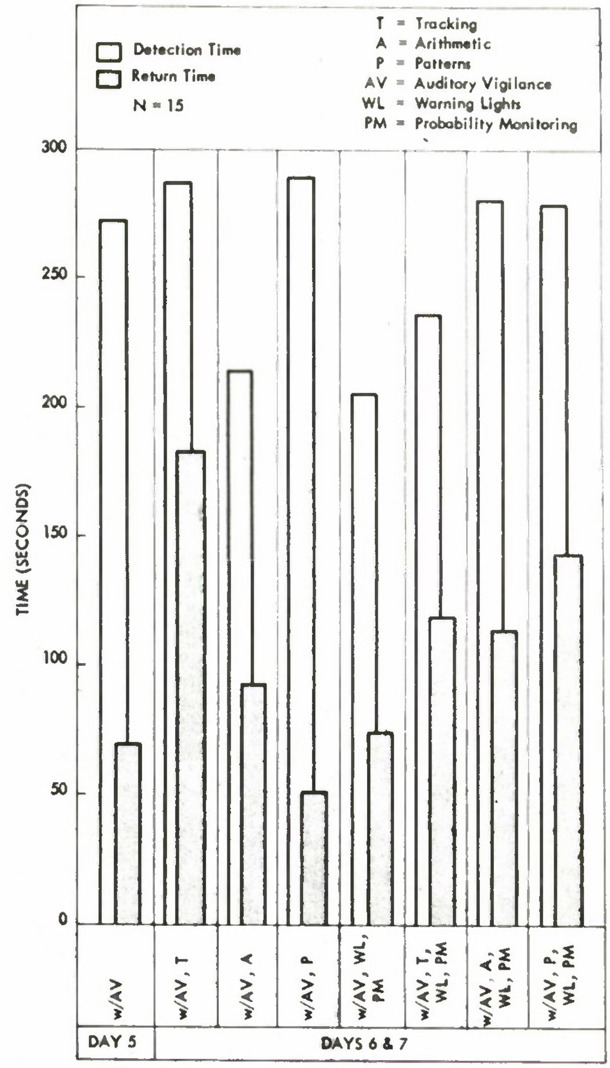


Figure 7. Comparison of Auditory Vigilance Scores (Number of False Responses + Number of Missed Signals) Obtained on Day 5 and During Performance with Other Tasks, As Indicated, on Days 6 and 7.

Figure 6 compares the means obtained during single- and multiple-task performance for the number of correct responses made on the pattern discrimination task. The differences in levels of performance are small, and the Friedman test shows them to be not significant.

The criterion measures obtained on the auditory vigilance task for the eight conditions of task performance are shown in Figures 7 and 8. Figure 7 contains a plot of the number of false responses, the number of missed signals, and the sum of these two, while Figure 8 contains a plot of the ratio of the number of correct responses to the number of signal changes plus the number of false responses.





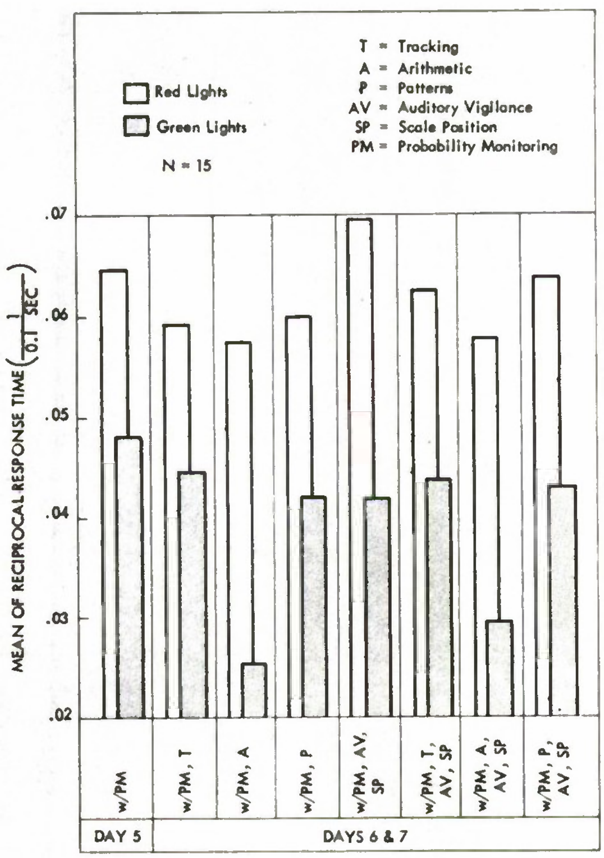


Figure 8 (ABOVE LEFT). Comparison of Auditory Vigilance Scores (Ratio of the Number of Correct Responses to the Number of Signal Changes Plus the Number of False Responses) Obtained on Day 5 and During Performance with Other Tasks, As Indicated, on Days 6 and 7.

Figure 9 (ABOVE RIGHT). Comparison of Scale Position Monitoring Scores Obtained on Day 5 and During Performance with Other Tasks, As Indicated, on Days 6 and 7.

Figure 10 (LEFT). Comparison of Warning Lights Scores Obtained on Day 5 and During Performance with Other Tasks, As Indicated, on Days 6 and 7.

The means of the detection time and of the return time for each of the eight conditions of performance on the scale position monitoring task are shown in Figure 9. Under each condition the mean detection time exceeds the mean return time, and both show rather pronounced changes with respect to several of the conditions of task performance. The statistical test shows that the differences among conditions are significant for both performance measures.

Figure 10 compares the mean reciprocal response times to green and to red warning lights during performance on Day 5 and during the seven conditions of multiple-task performance on Days 6 and 7. The response times to red warning lights are consistently shorter than those to green warning lights under all conditions of task performance. On the basis of Friedman's test, the hypothesis of no differences

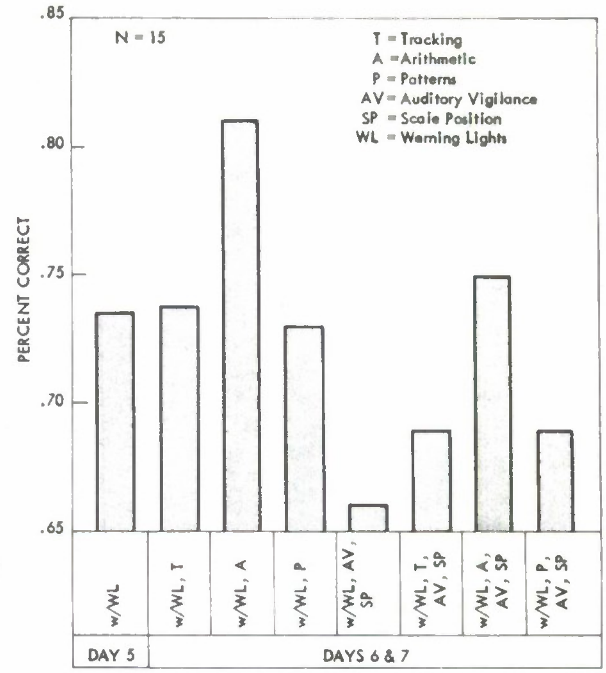


Figure 11. Comparison of Probability Monitoring Scores (Percentage of Correct Responses) Obtained on Day 5 and During Performance with Other Tasks, As Indicated, on Days 6 and 7.

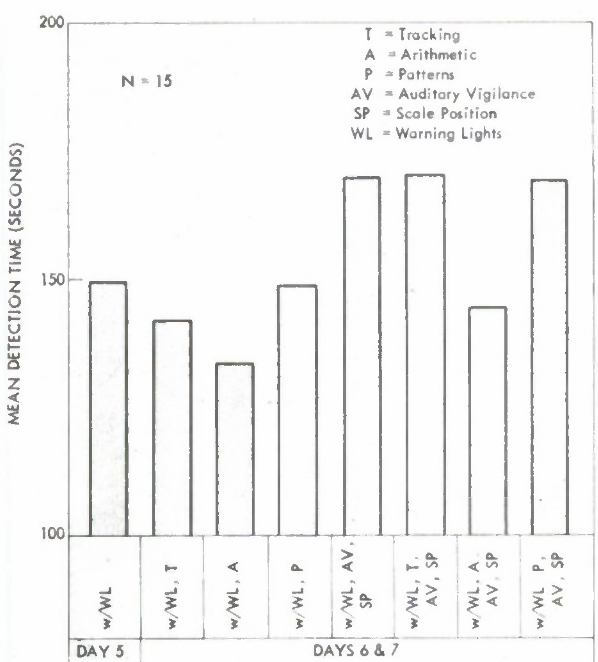


Figure 12. Comparison of Probability Monitoring Scores (Detection Time) Obtained on Day 5 and During Performance with Other Tasks, As Indicated, on Days 6 and 7.

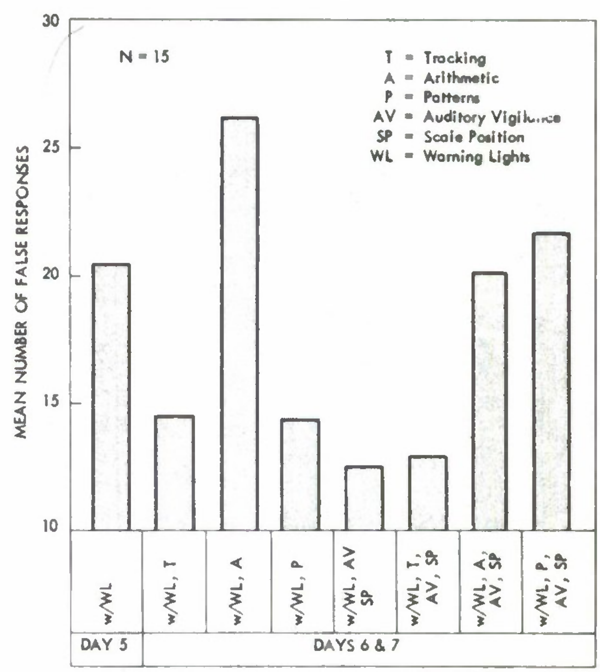


Figure 13. Comparison of Probability Monitoring Scores (Number of False Responses) Obtained on Day 5 and During Performance with Other Tasks, As Indicated, on Days 6 and 7.

among the conditions of task presentation can be rejected at less than the .001 level for the green-light response times and at the .02 level for the red-light response times.

The three criterion scores which were obtained during performance of the probability monitoring task are shown in Figures 11, 12 and 13. In Figure 11, a comparison is made of the mean percentage of correct responses for the various conditions of task presentation. Figure 12 shows the mean time required to detect a bias under each condition, and Figure 13 compares the mean number of false responses. The statistical test showed that the detection time and false response measures varied significantly as a function of changes in the task conditions.

Psychophysiological Measures

The psychophysiological data obtained prior to and during fifteen different performance task combinations on Days 5, 6, and 7 were used to evaluate

TABLE 5

Results of Friedman's Test of Significance of Differences Among Control and Among Activity Conditions for Each Psychophysiological Criterion Measure

Criterion Measure		df	χ_r^2	P-Level
Skin Resistance Level:	Control	14	15.33	--
	Activity	14	43.78	<.001
Skin Resistance Fluctuation:	Control	14	9.26	--
	Activity	14	59.37	<.001
Skin Temperature Level:	Control	14	1.59	--
	Activity	14	15.33	--
Skin Temperature Fluctuation:	Control	14	13.78	--
	Activity	14	30.13	<.01
Heart Rate Level:	Control	14	43.26	<.001
	Activity	14	56.65	<.001
Heart Rate Fluctuation:	Control	14	7.75	--
	Activity	14	21.06	.10
Respiration Rate Level:	Control	14	23.50	.05
	Activity	14	36.46	.001
Respiration Rate Fluctuation:	Control	14	9.45	--
	Activity	14	19.09	--

the effects on the autonomic substratum of variations in task combination. The scores for the psychophysiological measures taken during the control (00) intervals preceding the task presentation were ranked by subjects. The corresponding scores for the data from the activity intervals (01, 06, 12) were averaged and similarly ranked by subjects. Friedman's analysis of variance test for ranked data was applied to each of the measures. The Chi-squares yielded by this test are presented in Table 5.

The ranked data were also used for the computation of T-scores for each of the variables under each of the task combinations. These are presented in the following histograms (Figures 14 through 17). In each case, a higher T-score indicates a higher measure score.

Mean T-scores for the skin resistance level and fluctuation under various combinations of performance task presentation and for control conditions preceding the performance activity periods are shown in Figure 14. The Chi-square for the control period combinations did not reach significance for either level or fluctuation measures, but both the level and fluctuation measures showed significant Chi-squares (at the .001 level) for the activity intervals. The plots indicate more autonomic activity during the conditions in which the arith-

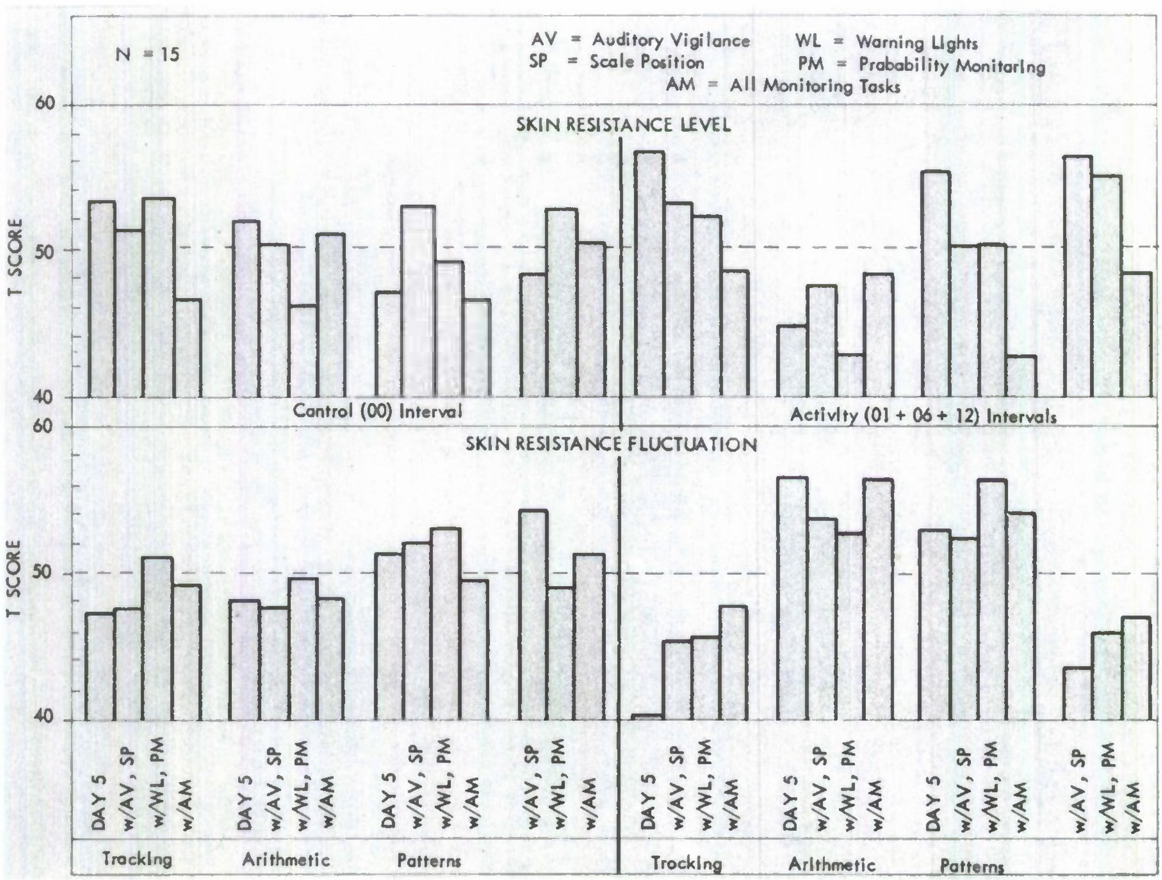


Figure 14. T-Scores for Skin Resistance Level and Fluctuation Measures Obtained During the Control and Activity Intervals for Each of the Task Conditions as Indicated.

metic and pattern discrimination tasks are performed than during the combinations involving tracking or only monitoring. During tracking, the histograms suggest that the autonomic activity increases with the addition of two monitoring tasks, and increases further with the addition of two more monitoring tasks.

Figure 15 shows similar plots of the skin temperature data. No significant Chi-squares were found except for the fluctuations associated with the activity intervals. The Friedman test showed that these latter differences are significant at the .01 level.

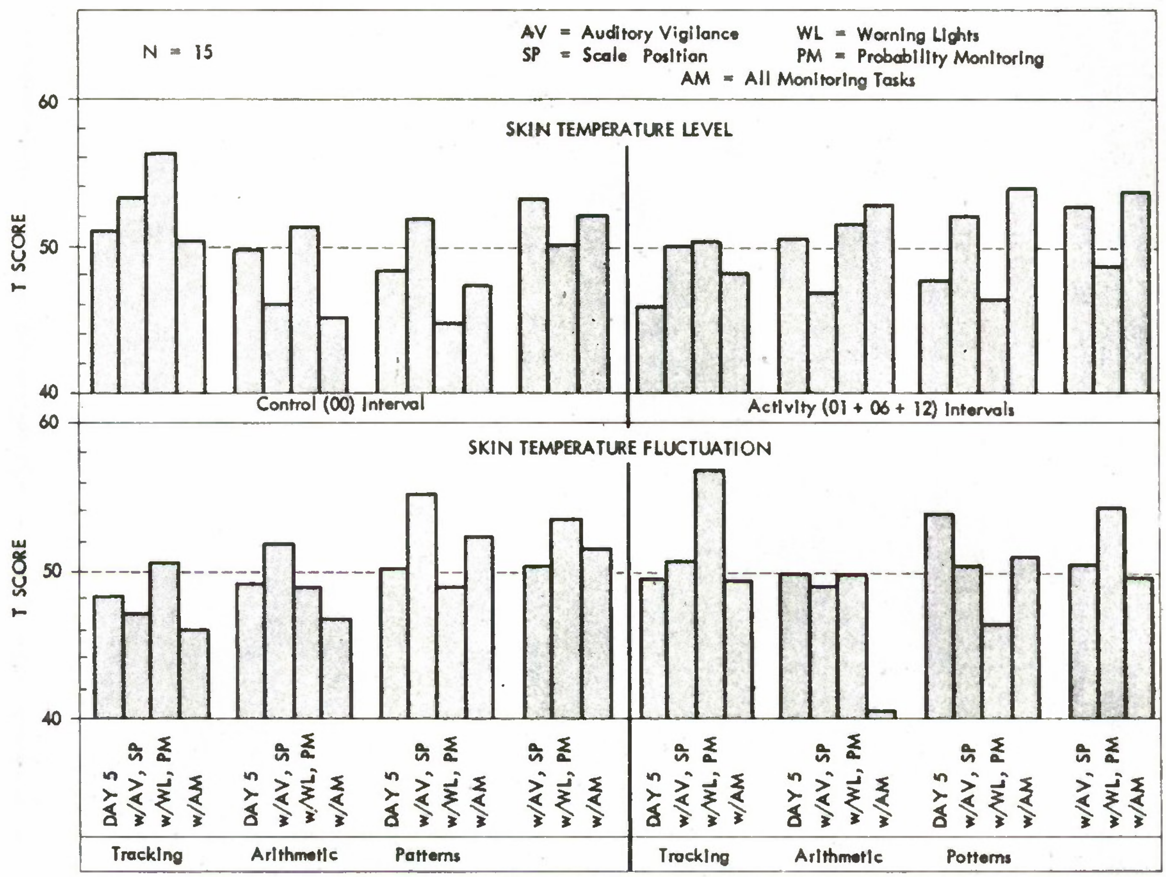


Figure 15. T-Scores for Skin Temperature Level and Fluctuation Measures Obtained During the Control and Activity Intervals for Each of the Task Conditions as Indicated.

The heart rate level and fluctuation T-scores are presented in Figure 16. The Friedman test shows significant differences ($P < .001$) in the scores for heart rate level, both in the activity intervals and in the control intervals. In both distributions, the heart rate level is considerably lower before and during the conditions involving the arithmetic task than with the other task combinations. No significant Chi-squares were found for the fluctuation measure.

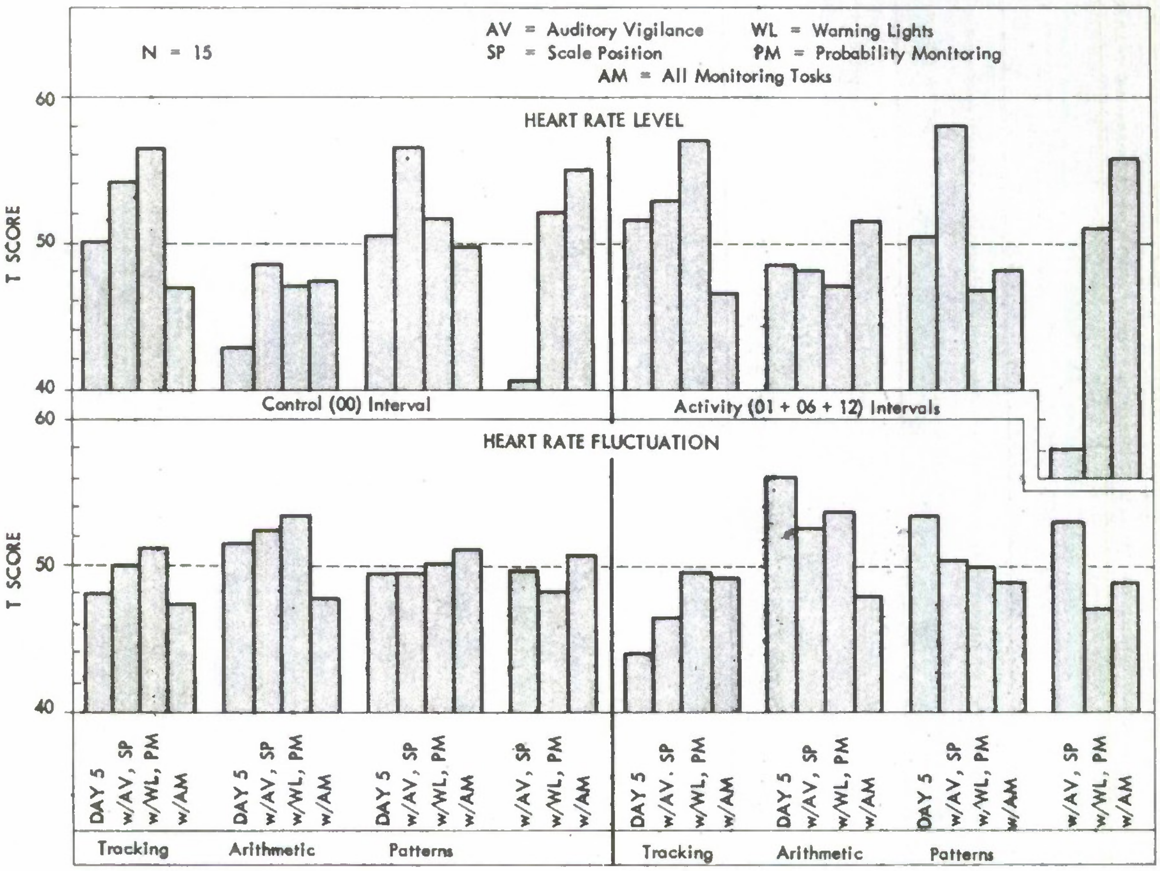


Figure 16. T-Scores for Heart Rate Level and Fluctuation Measures Obtained During the Control and Activity Intervals for Each of the Task Conditions as Indicated.

Plots of the respiration rate data are shown in Figure 17. Friedman's test showed the distribution of respiration rate level ranks during the activity intervals to exhibit differences significant at the .001 level. During the control interval the differences were significant at the .05 level. None of the Chi-squares for the other conditions plotted reached significance.

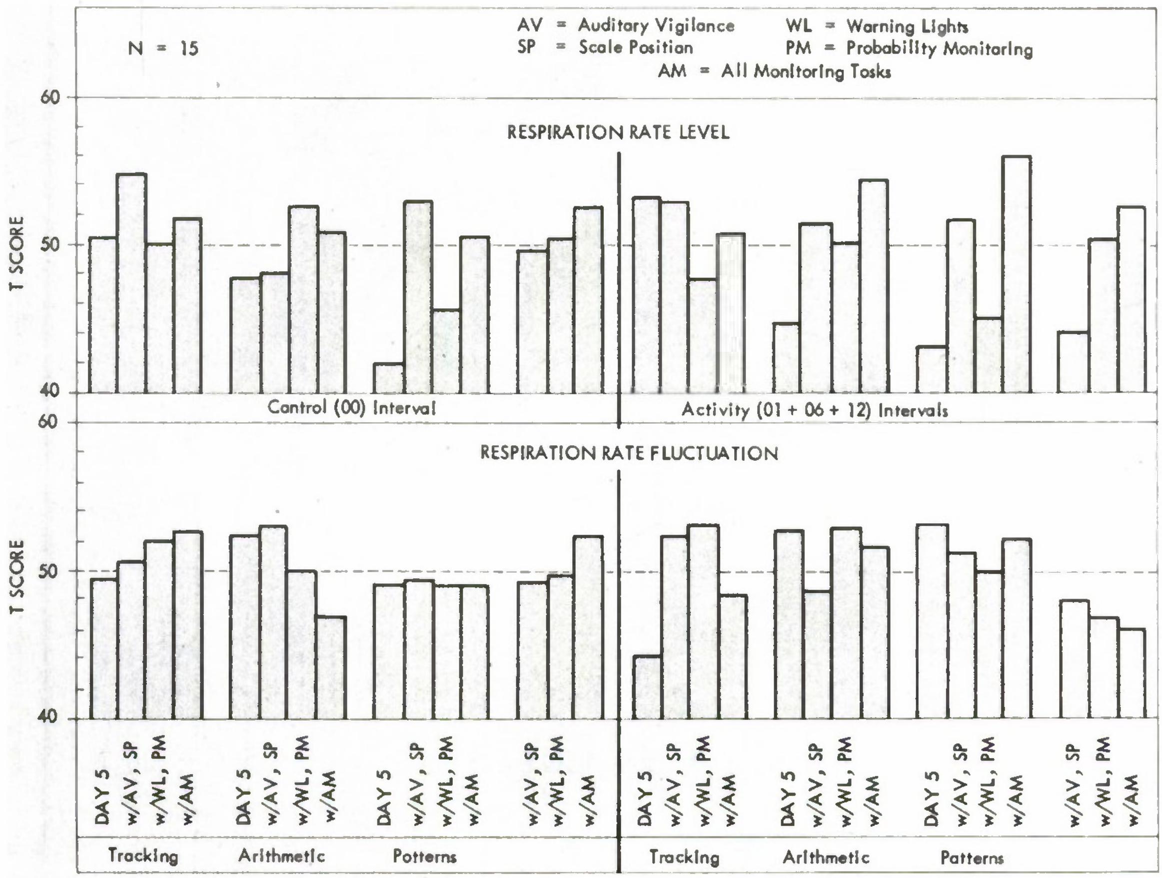


Figure 17. T-Scores for Respiration Rate Level and Fluctuation Measures Obtained During the Control and Activity Intervals for Each of the Task Conditions as Indicated.

DISCUSSION

With a few exceptions, the reliabilities of the performance criterion scores are encouraging, particularly in view of the fact that they are based on 24-hour test-retest measures. As has been mentioned previously, the tracking task scores might be questioned on the basis of the measuring apparatus, and several design changes are being studied with the intention of improving the reliabilities of the error sensing and recording circuits.

Two explanations are offered to account for the low reliabilities obtained for the auditory vigilance criterion measures. First, because the delay signals occurred with an average frequency of 4.4 times per hour, the subject's behavior was not adequately sampled during the course of a single hour. Second, although the delay signal exceeded the normal off signal by only 0.11 seconds (25%), it was apparently conspicuous enough to be readily discernable. It is believed that there is a combination of signal parameters and frequency of occurrence of the delay signal which would give better 24-hour reproducibility of scores.

The ease of the task may also be the best explanation for the low reliability of the return time measure for the scale position monitoring task. Namely, the subject knew that when he had observed and responded to an out-of-tolerance condition he could be certain that the scale would soon return to an in-tolerance limit, and he was therefore prepared for the occurrence of this event. As a result, the range of the return time scores was rather severely restricted, and such a restriction is known to lead to low correlation.

On the basis of the intercorrelations among the various tasks (Table 9, Appendix B), it appears that within the limits imposed by the sample size, these tasks measure independent psychomotor processes. However, a much more exhaustive study would have to be carried out to strengthen this conclusion.

No attempt has been made to derive a single score which would reflect the overall level of performance attained during each condition of task presentation. To do this would require assigning each task or criterion measure a position on a scale of task difficulty, and deriving a set of relative weights for each criterion measure. An alternative is to compare levels of performance when the scores are transformed to a common measure such as the T-score scale. Following this thought, each of the ranks which was assigned to the individual raw scores for the Friedman test analysis (Table 4) was transformed to a T-score. Two of the three probability monitoring task scores, detection time and percent correct, have been averaged since there was very little difference between their T-scores. The mean T-score computed for each condition of task performance was then taken as the best estimate of a given criterion score during a session. The resulting scores are shown in Figure 18. The plot in the upper-left panel represents the performance on Day 5 during single-task presentation while the remaining panels represent the ten conditions of multiple-task performance on Days 6 and 7. In each case, the T-score reflects goodness of performance.

Considering all criterion scores, it appears that the best performance was achieved during single-task presentation on Day 5 and during the condition in which the four monitoring tasks were presented simultaneously without the

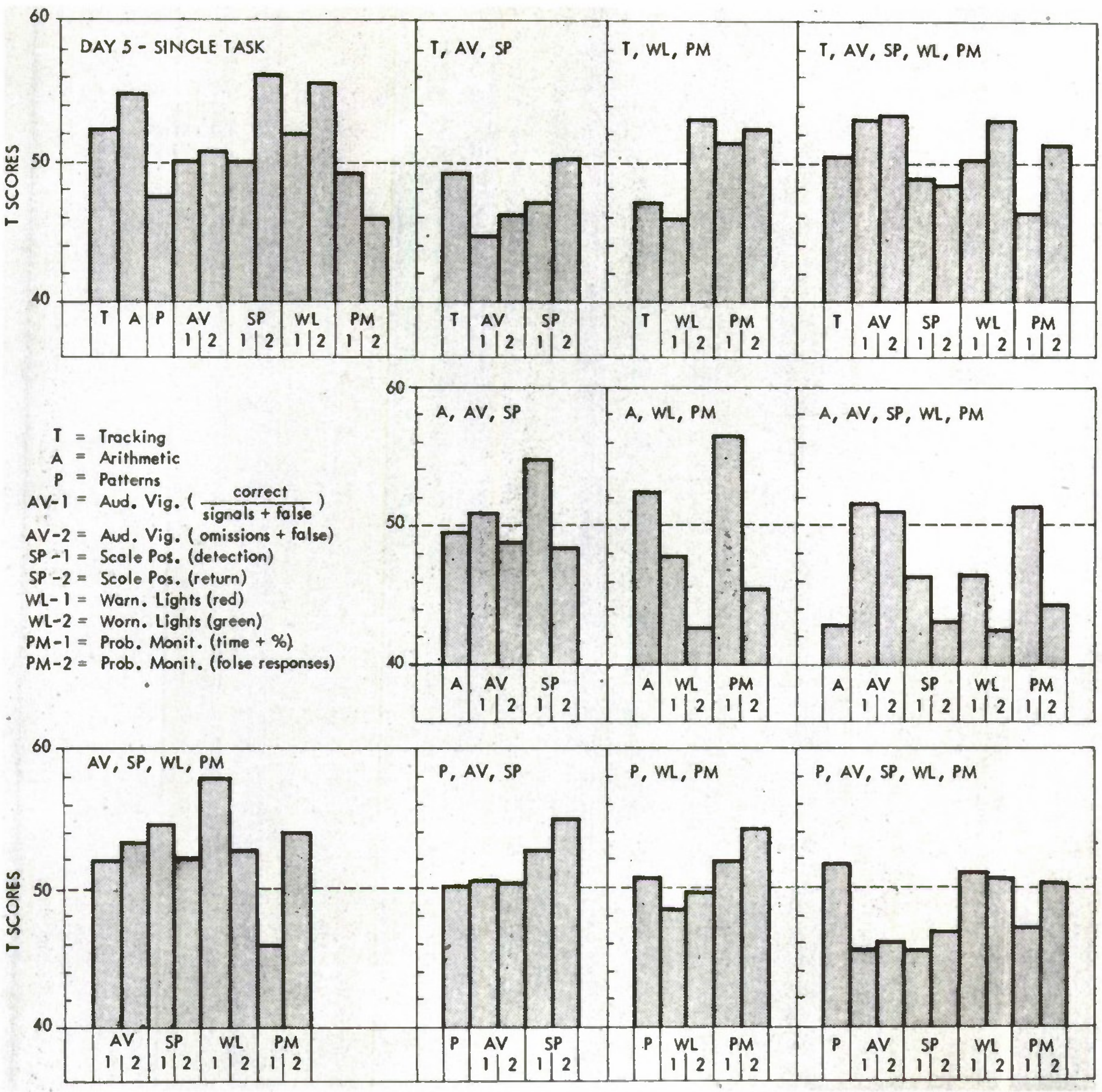


Figure 18. T-Scores for Each of the Performance Measures Obtained on Day 5 and During Multiple-Task Performance on Days 6 and 7.

tracking, arithmetic or pattern discrimination tasks (AV, SP, WL & PM). A fairly large decrement in performance is observed during the condition which combines the arithmetic task with all four monitoring tasks. While not as pronounced, there is an observable decrement during the condition of presentation of the pattern discrimination task with all four monitoring tasks.

Six of the nine criterion measures for the four monitoring tasks show a significant difference in levels of performance for the conditions of task presentation (see Table 4). To determine whether these six measures vary in the same manner, the eight conditions of monitoring task performance may be arbitrarily grouped into three categories. The first consists of the condition of single-task performance on Day 5, and of the condition which combines the four tasks (AV, SP, WL & PM). The second category consists of the six conditions which combine tracking, arithmetic or patterns with a single pair of monitoring tasks (T, AV & SP; T, WL & PM; A, AV & SP; A, WL & PM; P, AV & SP; and P, WL & PM). The third category consists of the combinations of tracking, arithmetic, and patterns with both pairs of monitoring tasks (T, AV, SP, WL & PM; A, AV, SP, WL & PM; and P, AV, SP, WL & PM). When the 48 T-scores are divided equally into 24 "high" scores and 24 "low" scores, the difference between the observed and expected frequencies (see Table 11, Appendix C) for the three categories is significant at the .05 level (Chi-square = 6.76). The difference is due to a greater than chance expectancy of high scores occurring in the first category, and of low scores occurring in the third category.

The psychophysiological studies were undertaken with several questions in mind. First, which of the measures, if any, maintain a moderately high reliability under conditions of rest? Second, which show reliability during task-performance activity? Third, which, if any, indicate level of performance of any given tasks? Fourth, do any of the measures respond similarly to substantially the same circumstances--that is, are any of the measures redundant? And fifth, which, if any, discriminate among tasks or task-combinations?

The rest- or control-period reliabilities of the measures were assessed on two bases: the reliability coefficients for Day 4 versus Day 5, and the similar coefficients for Day 6 versus Day 7. These were presented in Table 2 which indicates that all of the level measures except the heart-rate level were significantly correlated over days under both sets of circumstances as was the skin resistance fluctuation measure. It is interesting to compare the rank-order reliability coefficients obtained by Lacey and Lacey (19) for skin resistance fluctuation with those obtained in this study. These authors found a 48-hour reproducibility of .76; in the present study the corresponding 24-hour reproducibility was .86 for Day 4 versus Day 5 and .80 for Day 6 versus Day 7. The Laceys, however, also obtained a significant 48-hour reproducibility of .58 for heart rate fluctuation, whereas in the present study the coefficients were .32 and .07 for the pairs of days studied. The disagreement in the latter measure probably stems from differences in scoring methods used by the two groups of investigators.

In assessing the reliability of the psychophysiological measures during task-performance activity, data from the single-task performance sessions of Days 4 and 5 were compared and the reliability coefficients presented in Table 8, Appendix B. An examination of this table shows that none of the measures was significantly correlated under all of the single task conditions, but that skin resistance level and heart rate level showed correlations at the 5% level or better during four of the five task conditions; and respiration rate level and fluctuation showed such correlations during three of the five activity situations. Of all the measures, skin resistance level comes closest to exhibiting reliability under all the conditions of rest and activity.

The third question, whether any of the psychophysiological measures correlate with the level of performance of any of the activity tasks, was answered, essentially, by the observation that the 96 cross correlation coefficients between psychophysiological measure scores and performance task scores included only four at a significant level, and that this number is attributable to chance.

The question of whether any of the psychophysiological measures were redundant was examined by referring to the intercorrelation coefficients among these measures. (Table 10, Appendix B). All of the fluctuation measures were significantly correlated with each other, but the degree of correlation was not so high as to justify discarding any measure on the grounds of adding no useful information.

Several computations bear on the fifth question: whether there are any significant relationships among the psychophysiological measures and the presentation of performance tasks. The Chi-squares in Table 5, attained significant levels, .01 or better, for the following psychophysiological measures during activity intervals: respiration rate level, skin resistance level and fluctuation, skin temperature fluctuation, and heart rate level. In addition, heart rate fluctuation attained a significance level of .10. Two of the measures also indicated significant differences during the control intervals as a function of tasks to be presented during the following activity intervals. These were respiration rate level at the .05 level of significance, and heart rate level at the .001 level.

In examining the meanings of these Chi-squares, one first looks askance at the measures which show significance during the control periods. In the heart rate level T-score histograms, Figure 16, it is clear that a large contribution to the Chi-square is made, both for the control and for the activity intervals, by the very low score attained prior to and during the auditory vigilance and scale position tasks presented alone, and by the high score attained prior to and during the presentation of all monitoring tasks. Reference to the task scheduling charts in Tables 6 and 7, Appendix A, shows that the auditory vigilance and scale position tasks were scheduled during period 2 of Day 5 for each group, whereas all of the monitoring tasks were presented during periods 1 and 3 for the three groups. Diurnal variations of heart rate level were computed using data for all groups and all tasks, and resulted in the curve shown in Figure 19. From this it may be seen that the diurnal variations may well account for much of the difference in the two sets of scores, and no obvious conclusions should be drawn from the Chi-square scores for this measure. Similar considerations apply to the respiration rate level measures and to the skin temperature fluctuation measure.

However, diurnal variations in skin resistance fluctuation were essentially cancelled by the program of scheduling tasks, insofar as contribution to significant Chi-square is concerned, and the variations seen in Figure 14 may be taken to represent a fair picture of the psychophysiological situation. It seems clear that the highly significant Chi-square for skin resistance fluctuation arises from higher than average fluctuations during arithmetic and patterns tasks and lower than average fluctuations during tracking and monitoring tasks. A somewhat similar pattern is seen in the histograms for heart rate fluctuations during activity intervals, although the Chi-square

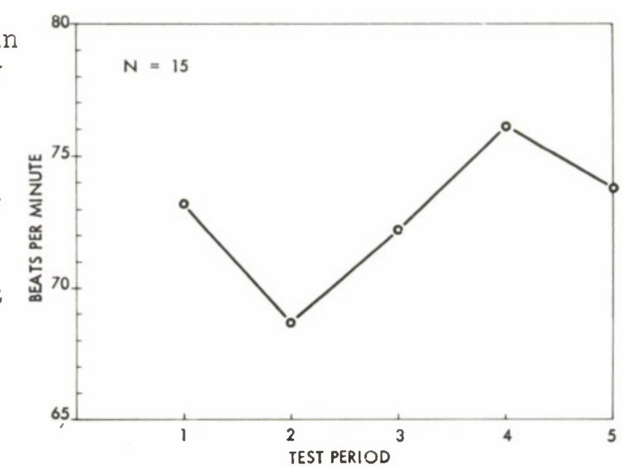


Figure 19. Diurnal Variations in Heart Rate Level.

for this distribution reaches only the .10 level of significance. The skin resistance level during activity intervals shows a considerably lower average skin resistance during arithmetic tasks than during others, so that at least three of the measures indicate higher autonomic activity during the arithmetic tasks.

Each of the fluctuation measures was found to have some significant degree of correlation with each of the others. To determine whether the fluctuation measures, in general, vary with the task conditions, the 60 T-scores obtained for these four measures under the 15 task combinations were divided equally into "high" and "low" scores. When the task conditions are grouped into the four categories shown in Figures 14 through 17 (tracking combinations, arithmetic combinations, pattern combinations, and monitoring alone), the difference between the observed and expected frequencies (see Table 12, Appendix C) of "high" and "low" scores is significant at the .005 level (Chi-square = 14.25). The frequency of "high" scores occurring during conditions of arithmetic and of pattern task performance, and of "low" scores during conditions of tracking and of monitoring task performance, is greater than would be expected by chance.

SUMMARY

Fifteen male subjects were studied in an experiment designed to investigate six problems: (a) the 24-hour test-retest reliability of a predetermined battery of seven performance tasks yielding 12 criterion scores; (b) the degrees of intercorrelation among the 12 performance criterion scores; (c) the effects of performing simultaneously various combinations of the tasks which are physically compatible; (d) the 24-hour test-retest reliability of a predetermined battery of four psychophysiological variables yielding eight criterion measures; (e) the degrees of intercorrelation among the eight psychophysiological criterion measures; and (f) the relationships of individual psychophysiological criterion measures to performance task criterion scores, and to conditions of task performance. The performance task designations and the number of criterion scores yielded by each were as follows: compensatory tracking (1), arithmetic computation (1), pattern discrimination (1), auditory vigilance (2), scale position monitoring (2), warning lights monitoring (2), and probability monitoring (3). The psychophysiological variables were as follows: skin resistance, skin temperature, heart rate, and respiration rate. Each of these variables yielded two criterion measures corresponding to the mean level and to the fluctuation of the variable.

The results are necessarily tentative since they are based on a small sample of subjects, and were derived from an experimental design which did not permit the control of some systematic error effects. With these reservations, the findings were as follows:

1. Criterion scores for five tasks demonstrated moderately high reliability considering that 24-hour test-retest reliability is rarely outstanding for most performance tasks. These were arithmetic computation, pattern discrimination, scale position monitoring, warning lights monitoring, and probability monitoring.

2. With possibly two exceptions, one involving warning lights monitoring and the other involving probability monitoring, none of the intercorrelations obtained for the performance criterion scores can be considered significant.

3. Seven of the performance criterion measures reflected a significant difference among the several conditions in which a varying number of tasks were performed. These involved arithmetic computation, scale position monitoring, warning lights monitoring and probability monitoring. When all tasks and all conditions of task presentation were considered, a decrease in criterion scores was observed as the number of concurrently performed tasks increased.

4. Three of the eight psychophysiological measures showed fairly high 24-hour test-retest reliability for the control or resting conditions. These were skin temperature level, skin resistance fluctuation, and respiration rate level. For several different task-activity conditions moderate 24-hour test-retest reliability was shown by the skin resistance level, the heart rate level, and the respiration rate level measures.

5. All of the psychophysiological fluctuation measures were significantly intercorrelated, during resting conditions. None of the level measures showed significant correlations either with any other level measures or with any of the fluctuation measures.

6. The number of significant intercorrelations among the eight psychophysiological criterion measures and the 12 performance criterion scores was so small as to be attributable to chance. Several of the psychophysiological measures, however, showed differences with different tasks or task conditions.

REFERENCES

1. Adams, O. S. Aircrew fatigue problems during extended endurance flight, Phase I: Planning, WADC Technical Report 57-510, May 1958.

2. Ax, A. F. The physiological differentiation between fear and anger in humans. Psychosom. Med. 1953, 15, 433-442.

3. Baker, K. E., Wylie, R. C., and Gagne, R. M. The effects of an interfering task on the learning of a complex motor skill. J. Exp. Psychol., 1951, 41, 1-9.

4. Baker, L. M., and Taylor, W. M. The relationship under stress between changes in skin temperature, electrical skin resistance, and pulse rate. J. Exp. Psychol., 1954, 48, 361-366.

5. Bartlett, F. C. Fatigue following highly skilled work. Proc. Roy. Soc., Series B, 1943, 131, 247-257.

6. Benedict, R. B. Effect of conscious anxiety on the electrocardiogram. J. Amer. Med. Women's Assoc., 1951, 6, 419-422.

7. Blatz, W. E. The cardiac, respiratory, and electrical phenomena involved in the emotion of fear. J. Exp. Psychol., 1925, 8, 109-132.

8. Conrad, R. Some effects on performance of changes in perceptual load. J. Exp. Psychol., 1955, 49, 313-322.

9. Corey, E. L. Pilot metabolism and respiratory activity during varied flight tasks. J. Appl. Physiol., 1948, 1, 35-44.

10. Craik, K. J. W. Fatigue apparatus, Flying Personnel Research Committee Report No. 119, British Air Ministry, 1940.

11. Dixon, W. J., and Massey, F. J. Introduction to statistical analysis. New York: McGraw-Hill, 1951.

12. Ford, A. Bioelectrical potentials and mental effort: I. Cardiac effects. J. Comp. Physiol. Psychol., 1953, 46, 347-351.

13. Friedman, M. The use of ranks to avoid the assumption of normality. J. Amer. Stat. Assoc., 1937, 32, 675-701.

14. Haggard, E. A. On the application of analysis of variance to GSR data: II. Some effects of the use of inappropriate measures. J. Exper. Psychol., 1949, 39, 861-867.

15. Hartley, H. O. The maximum F-ratio as a short-cut test for heterogeneity of variance. Biometrika, 1950, 37, 308-312.

16. Hauty, G. T., and Payne, R. B. Effects of work prolongation upon components of a perceptual-motor task, USAF School of Aviation Medicine Project No. 21-1601-004, Report No. 6, October 1954.

17. Howland, D. An investigation of the performance of the human monitor, WADC Technical Note 57-431, July 1958.

18. Lacey, J. I. Individual differences in somatic response patterns. J. comp. physiol. Psychol., 1950, 43, 338-350.

19. Lacey, J. I., and Lacey, B. C. The relationship of resting autonomic activity to motor impulsivity. In Proceedings of the Association for Research in Nervous and Mental Disease, Vol. 36, The Brain and Human Behavior. Baltimore: Williams & Wilkins, 1958.

20. Melton, A. W. (Ed.) Apparatus tests, AAF Aviat. Psychol. Program, Report No. 4, 1947.

21. Notterman, J. M. Experimental anxiety and a conditioned heart rate response in human beings. Trans. N. Y. Acad. Sci., 1953, 16, 24-33.

22. Scholsberg, H., and Stanley, W. C. A simple test of the normality of twenty-four hour distributions of electrical skin conductance. Science, 1953, 117, 35-36.

23. Staudt, V., and Kubis, J. F. The psychogalvanic response (PGR) and its relation to changes in tension and relaxation. J. Psychol., 1948, 25, 443-453.

24. van der Valk, J. M., and Groen J. Electrical resistance of the skin during induced emotional stress. Psychosom. Med., 1950, 12, 303-314.

25. Woodworth, R. S., and Schlosberg, H. Experimental Psychology. (Rev. ed.) New York: Holt, 1954.

APPENDIX A

TASK PRESENTATION SCHEDULES FOR PHASES I AND II

The task presentation schedules for each of the three groups during Phase I (Days 1-5), and during Phase II (Days 6 and 7), are shown in Tables 6 and 7, respectively. The tasks have been assigned the following letter designations:

A - Compensatory Tracking
B - Arithmetic Computation
C - Pattern Discrimination
D - Auditory Vigilance + Scale Position Monitoring
E - Warning Lights Monitoring + Probability Monitoring

TABLE 6

Task Presentation Schedule for Each Group for the Five Hours of Daily Testing During Phase I

		Group I					Group II					Group III				
		Period					Period					Period				
		1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Learning	Day 1	A	D	B	E	C	C	D	A	E	B	B	D	C	E	A
	Day 2	B	E	C	D	A	A	E	B	D	C	C	E	A	D	B
	Day 3	C	D	A	E	B	B	D	C	E	A	A	D	B	E	C
Reliability	Day 4	A	E	C	D	B	C	E	B	D	A	B	E	A	D	C
	Day 5	B	D	A	E	C	A	D	C	E	B	C	D	B	E	A

TABLE 7

Task Presentation Schedule for Each Group for the Five Hours of Daily Testing During Phase II

	Group I					Group II					Group III				
	Period					Period					Period				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Day 6	DE	BD	ADE	CE	AD	CE	BDE	AD	CDE	BE	CD	BE	CDE	AE	DE
Day 7	CD	BDE	AE	CDE	BE	AE	BD	ADE	CD	DE	BD	ADE	CE	BDE	AD

APPENDIX B

TABLE 8

Reliability Coefficients for Psychophysiological Measures During Performance as Specified (Day 4 vs. Day 5)

Measure	Track.	Arith.	Patt.	AV-SP	WL-PM
Skin Resistance					
Level	.732**	.089	.668**	.798**	.781**
Fluctuation	.796**	.421	.415	.623*	.320
Skin Temperature					
Level	.490	.285	.890**	.031	.438
Fluctuation	.268	-.066	-.037	.071	.069
Heart Rate					
Level	.533*	.823**	.210	.662**	.561*
Fluctuation	.419	-.005	.354	.524*	.315
Respiration Rate					
Level	.742**	.358	.435	.576*	.753**
Fluctuation	.673**	.643**	-.133	.538*	.474

* $P \leq .05$
** $P \leq .01$

TABLE 9

Intercorrelation of Performance Measures

Measure	1	2	3	4	5	6	7	8	9	10	11	12
1. Tracking (H + V RMS Error)		-.155	-.414	-.461	.529	.565	.429	-.049	.130	.284	-.344	.357
2. Arithmetic (Number Correct)			.140	-.323	.280	.229	.268	-.257	-.307	-.187	.031	.111
3. Pattern Discrimination (Number Correct)				-.237	.190	.023	.325	.257	-.130	.341	-.340	.063
4. Auditory Vigilance (No. Correct/No. Signals + False)					-.964**	-.338	-.225	-.005	.321	-.134	.158	-.407
5. Auditory Vigilance (No. False + No. Missed)						.324	.281	.071	-.239	.252	-.263	.485
6. Scale Position Monitoring (Detection Time							.022	-.161	-.319	-.051	.109	.000
7. Scale Position Monitoring (Return Time)								.281	.189	.160	-.317	.241
8. Warning Lights Monitoring (Red-Light Response Time)									.750**	.182	-.201	-.002
9. Warning Lights Monitoring (Green-Light Response Time)										-.078	-.066	-.042
10. Probability Monitoring (No. Correct/No. Biases)											-.191	.498
11. Probability Monitoring (Detection Time)												-.711**
12. Probability Monitoring (No. False Responses)												

** P < .01

TABLE 10

Intercorrelation Coefficients for Psychophysiological Measures on Days 4, 5, 6, and 7 During Control (00) Intervals

Measure	1	2	3	4	5	6	7	8
1. Skin Resistance Level		-.371	.215	.464	-.214	.096	.185	.007
2. Skin Resistance Fluctuation			.228	.544*	-.090	.588*	.380	.774**
3. Skin Temperature Level				.258	.350	-.103	.403	.053
4. Skin Temperature Fluctuation					-.308	.621**	.318	.773**
5. Heart Rate Level						-.293	.166	-.335
6. Heart Rate Fluctuation							-.203	.913**
7. Respiration Rate Level								-.045
8. Respiration Rate Fluctuation								

* $P < .05$
** $P \leq .01$

APPENDIX C

TABLE 11

Comparison of the Observed vs. Expected Frequencies of "High" and "Low" T-Scores for the Scale Position, Warning Lights, and Probability Monitoring Tasks with Respect to the Three Categories of Task Conditions as Shown*

Score		Day 5 AV,SP,WL & PM	T, A, P and AV-SP or WL-PM	T,AV,SP,WL & PM A,AV,SP,WL & PM P,AV,SP,WL & PM
High:	Observed	9	10	5
	(Expected)	(6)	(9)	(9)
Low:	Observed	3	8	13
	(Expected)	(6)	(9)	(9)

* Chi-square = 6.76 ($P < .05$)

TABLE 12

Comparison of the Observed vs. Expected Frequencies of "High" and "Low" T-Scores for the Psychophysiological Fluctuation Measures with Respect to the Four Categories of Task Conditions as Shown*

Score		T (Day 5) T,AV-SP T,WL-PM T,AV-SP,WL-PM	A (Day 5) A,AV-SP A,WL-PM A,AV-SP,WL-PM	P (Day 5) P,AV-SP P,WL-PM P,AV-SP,WL-PM	AV-SP (Day 5) WL-PM (Day 5) AV-SP, WL-PM
High:	Observed	4	10	13	3
	(Expected)	(8)	(8)	(8)	(6)
Low:	Observed	12	6	3	9
	(Expected)	(8)	(8)	(8)	(6)

* Chi-square = 14.25 ($P < .005$)

AD- 213592

Wright Air Development Center, Aero Medical Laboratory, Wright-Patterson Air Force Base, Ohio; Lockheed Aircraft Corporation, Marietta, Georgia
RESEARCH TO INVESTIGATE FACTORS AFFECTING MULTIPLE-TASK PSYCHOMOTOR PERFORMANCE, by Oscar S. Adams, Raphael B. Levine and W. Dean Chiles, March 1959, 37 p. incl. illus. tables, 25 ref. (Proj 7184; Task 71582)
(WADC TR 59-120)
(Contract AF 33(616)-6050
Unclassified report

This study showed the performance tests used to have moderately high reliability and sensitivity to work load but a negligible

(over)

UNCLASSIFIED

1. Psychology
2. Psychophysiology
3. Task Load
I. Adams, Oscar S.
II. Levine, R.
III. Chiles, W. Dean

UNCLASSIFIED

AD- 213592

Wright Air Development Center, Aero Medical Laboratory, Wright-Patterson Air Force Base, Ohio; Lockheed Aircraft Corporation, Marietta, Georgia
RESEARCH TO INVESTIGATE FACTORS AFFECTING MULTIPLE-TASK PSYCHOMOTOR PERFORMANCE, by Oscar S. Adams, Raphael B. Levine and W. Dean Chiles, March 1959, 37 p. incl. illus. tables, 25 ref. (Proj 7184; Task 71582)
(WADC TR 59-120)
(Contract AF 33(616)-6050
Unclassified report

This study showed the performance tests used to have moderately high reliability and sensitivity to work load but a negligible

(over)

UNCLASSIFIED

1. Psychology
2. Psychophysiology
3. Task Load
I. Adams, Oscar S.
II. Levine, R.
III. Chiles, W. Dean

UNCLASSIFIED

AD- 213592

amount of intercorrelation. While no significant correlation was obtained between performance scores and measures of psychophysiological activity, several of the psychophysiological measures showed significant reliability, intercorrelation, and sensitivity to conditions of rest and performance activity.

UNCLASSIFIED

UNCLASSIFIED

AD- 213592

amount of intercorrelation. While no significant correlation was obtained between performance scores and measures of psychophysiological activity, several of the psychophysiological measures showed significant reliability, intercorrelation, and sensitivity to conditions of rest and performance activity.

UNCLASSIFIED

UNCLASSIFIED

AD- 213592

Wright Air Development Center, Aero Medical Laboratory, Wright-Patterson Air Force Base, Ohio; Lockheed Aircraft Corporation, Marietta, Georgia
RESEARCH TO INVESTIGATE FACTORS AFFECTING MULTIPLE-TASK PSYCHOMOTOR PERFORMANCE, by Oscar S. Adams, Raphael B. Levine and W. Dean Chiles, March 1959, 37 p. incl. illus. tables, 25 ref. (Proj 7184; Task 71582)
(WADC TR 59-120)
(Contract AF 33(616)-6050

Unclassified report

This study showed the performance tests used to have moderately high reliability and sensitivity to work load but a negligible

(over)

UNCLASSIFIED

1. Psychology
2. Psychophysiology
3. Task Load
I. Adams, Oscar S.
II. Levine, R.
III. Chiles, W. Dean

UNCLASSIFIED

AD- 213592

Wright Air Development Center, Aero Medical Laboratory, Wright-Patterson Air Force Base, Ohio; Lockheed Aircraft Corporation, Marietta, Georgia
RESEARCH TO INVESTIGATE FACTORS AFFECTING MULTIPLE-TASK PSYCHOMOTOR PERFORMANCE, by Oscar S. Adams, Raphael B. Levine and W. Dean Chiles, March 1959, 37 p. incl. illus. tables, 25 ref. (Proj 7184; Task 71582)
(WADC TR 59-120)
(Contract AF 33(616)-6050

Unclassified report

This study showed the performance tests used to have moderately high reliability and sensitivity to work load but a negligible

(over)

UNCLASSIFIED

1. Psychology
2. Psychophysiology
3. Task Load
I. Adams, Oscar S.
II. Levine, R.
III. Chiles, W. Dean

UNCLASSIFIED

AD- 213592

amount of intercorrelation. While no significant correlation was obtained between performance scores and measures of psychophysiological activity, several of the psychophysiological measures showed significant reliability, intercorrelation, and sensitivity to conditions of rest and performance activity.

UNCLASSIFIED

UNCLASSIFIED

AD- 213592

amount of intercorrelation. While no significant correlation was obtained between performance scores and measures of psychophysiological activity, several of the psychophysiological measures showed significant reliability, intercorrelation, and sensitivity to conditions of rest and performance activity.

UNCLASSIFIED

UNCLASSIFIED

AD- 213592

Wright Air Development Center, Aero Medical Laboratory, Wright-Patterson Air Force Base, Ohio; Lockheed Aircraft Corporation, Marietta, Georgia
RESEARCH TO INVESTIGATE FACTORS AFFECTING MULTIPLE-TASK PSYCHOMOTOR PERFORMANCE, by Oscar S. Adams, Raphael B. Levine and W. Dean Chiles, March 1959, 37 p. incl. illus. tables, 25 ref. (Proj 7184; Task 71582)
(WADC TR 59-120)
(Contract AF 33(616)-6050
Unclassified report

This study showed the performance tests used to have moderately high reliability and sensitivity to work load but a negligible

(over)

UNCLASSIFIED

1. Psychology
2. Psychophysiology
3. Task Load
I. Adams, Oscar S.
II. Levine, R.
III. Chiles, W. Dean

UNCLASSIFIED

AD- 213592

Wright Air Development Center, Aero Medical Laboratory, Wright-Patterson Air Force Base, Ohio; Lockheed Aircraft Corporation, Marietta, Georgia
RESEARCH TO INVESTIGATE FACTORS AFFECTING MULTIPLE-TASK PSYCHOMOTOR PERFORMANCE, by Oscar S. Adams, Raphael B. Levine and W. Dean Chiles, March 1959, 37 p. incl. illus. tables, 25 ref. (Proj 7184; Task 71582)
(WADC TR 59-120)
(Contract AF 33(616)-6050
Unclassified report

This study showed the performance tests used to have moderately high reliability and sensitivity to work load but a negligible

(over)

UNCLASSIFIED

1. Psychology
2. Psychophysiology
3. Task Load
I. Adams, Oscar S.
II. Levine, R.
III. Chiles, W. Dean

UNCLASSIFIED

AD- 213592

amount of intercorrelation. While no significant correlation was obtained between performance scores and measures of psychophysiological activity, several of the psychophysiological measures showed significant reliability, intercorrelation, and sensitivity to conditions of rest and performance activity.

UNCLASSIFIED

UNCLASSIFIED

AD- 213592

amount of intercorrelation. While no significant correlation was obtained between performance scores and measures of psychophysiological activity, several of the psychophysiological measures showed significant reliability, intercorrelation, and sensitivity to conditions of rest and performance activity.

UNCLASSIFIED

UNCLASSIFIED

AD- 213592

Wright Air Development Center, Aero Medical Laboratory, Wright-Patterson Air Force Base, Ohio; Lockheed Aircraft Corporation, Marietta, Georgia
RESEARCH TO INVESTIGATE FACTORS AFFECTING MULTIPLE-TASK PSYCHOMOTOR PERFORMANCE, by Oscar S. Adams, Raphael B. Levine and W. Dean Chiles, March 1959, 37 p. incl. illus. tables, 25 ref. (Proj 7184; Task 71582)
(WADC TR 59-120)
(Contract AF 33(616)-6050
Unclassified report

This study showed the performance tests used to have moderately high reliability and sensitivity to work load but a negligible

(over)

UNCLASSIFIED

1. Psychology
2. Psychophysiology
3. Task Load
I. Adams, Oscar S.
II. Levine, R.
III. Chiles, W. Dean

UNCLASSIFIED

AD- 213592

Wright Air Development Center, Aero Medical Laboratory, Wright-Patterson Air Force Base, Ohio; Lockheed Aircraft Corporation, Marietta, Georgia
RESEARCH TO INVESTIGATE FACTORS AFFECTING MULTIPLE-TASK PSYCHOMOTOR PERFORMANCE, by Oscar S. Adams, Raphael B. Levine and W. Dean Chiles, March 1959, 37 p. incl. illus. tables, 25 ref. (Proj 7184; Task 71582)
(WADC TR 59-120)
(Contract AF 33(616)-6050
Unclassified report

This study showed the performance tests used to have moderately high reliability and sensitivity to work load but a negligible

(over)

UNCLASSIFIED

1. Psychology
2. Psychophysiology
3. Task Load
I. Adams, Oscar S.
II. Levine, R.
III. Chiles, W. Dean

UNCLASSIFIED

AD- 213592

amount of intercorrelation. While no significant correlation was obtained between performance scores and measures of psychophysiological activity, several of the psychophysiological measures showed significant reliability, intercorrelation, and sensitivity to conditions of rest and performance activity.

UNCLASSIFIED

UNCLASSIFIED

AD- 213592

amount of intercorrelation. While no significant correlation was obtained between performance scores and measures of psychophysiological activity, several of the psychophysiological measures showed significant reliability, intercorrelation, and sensitivity to conditions of rest and performance activity.

UNCLASSIFIED

UNCLASSIFIED