AD-A061 725 FEDERAL AVIATION ADMINISTRATION OKLAHOMA CITY OKLA CI--ETC F/G 5/10
THE MEASUREMENT AND SCALING OF WORKLOAD IN COMPLEX PERFORMANCE,(U)
SEP 78 W D CHILES, A E JENNINGS, E A ALLUISI
UNCLASSIFIED FAA-AM-78-34 NL

1 OF 1
AD
A061725

END
DATE
FILMED
2-79
DDC

FAA-AM-78-34

THE MEASUREMENT AND SCALING OF WORKLOAD IN COMPLEX PERFORMANCE

W. Dean Chiles
Alan E. Jennings
Earl A. Alluisi*
FAA Civil Aeromedical Institute
P.O. Box 25082
Oklahoma City, Oklahoma 73125

*Performance Assessment Laboratory
Old Dominion University
Norfolk, Virginia 23508

SEPTEMBER 1978

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

Prepared for
U.S. DEPARTMENT OF TRANSPORTATION
Federal Aviation Administration
Office of Aviation Medicine
Washington, D.C. 20591

NOTICE

Technical Report Documentation Page

(18) 1. Report No. (19) FAA-AM-78-34	2. Government Accession No.	3. Recipient's Catalog No.
(6) 4. Title and Subtitle THE MEASUREMENT AND SCALING OF WORKLOAD IN COMPLEX PERFORMANCE		(11) 5. ~~Report Date~~ SEPTEMBER 1978 6. ~~Performing Organization Code~~
(10) 7. Author(s) W. Dean Chiles, Alan E. Jennings, and Earl A. Alluisi		8. Performing Organization Report No.
9. Performing Organization Name and Address FAA Civil Aeromedical Institute P.O. Box 25082 Oklahoma City, Oklahoma 73125		10. Work Unit No. (TRAIS) 11. Contract or Grant No. 13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address Office of Aviation Medicine Federal Aviation Administration 800 Independence Avenue, S.W. Washington, D.C. 20591		14. Sponsoring Agency Code

15. Supplementary Notes

Work was performed under Tasks AM-D-77/78-PSY-57 (12) 15 p.

16. Abstract

Two groups of young men (Group I, N = 51, tested identically on 2 successive days; Group II, N = 43, tested on 1 day only) performed various combinations of the six tasks of the CAMI Multiple Task Performance Battery. Two of the tasks involved the monitoring of static (lights) and dynamic (meters) processes; the four more-active tasks involved mental arithmetic, elementary problem solving, pattern identification, and two-dimensional compensatory tracking. Five of nine performance intervals provided different complex tasks consisting of both of the monitoring tasks and two of the active tasks presented concurrently. Other trials provided data on the singly performed constituent tasks as well as the combined monitoring tasks. Results indicated that all 12 performance measures varied significantly as a function of the different task-combination conditions. A standard psychological scaling technique (Thurstone Case V) was applied to the monitoring data (for the green and red lights and for the meters) to develop an index of workload for the five complex task combinations. Since better performance was presumed to indicate a lower workload, workload was inferred to increase as performance declined across conditions. The best performances (scale values of zero) were associated with single tasks as expected. Scale values for the complex task-combination conditions were consistent between groups and between the 2 days of testing of Group I (r's of .947 to .993). Although the scale values are specific to the tasks and task-combination conditions employed, the scaling-procedure application shows promise for cases in which quantitative measures of performance can be acquired with moderately large ($N \geq 50$) samples of subjects.

greater than

17. Key Words Scaling Workload Complex Performance	18. Distribution Statement Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 12	22. Price

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

084 050

THE MEASUREMENT AND SCALING OF WORKLOAD IN COMPLEX PERFORMANCE

I. Introduction.

The level of performance that can be expected of a properly trained and selected human operator is of great interest in many decisions involving the design of man-machine systems. In many transportation areas, the level of operator performance has direct implications for safety. The level of performance has generally been held to be a function of three broad categories of influences: personnel factors, situational factors, and job demands. Job demands, i.e. workload, can be considered to be related to the number and variety of skills that the operator must exercise in performing his job and to the nature of the specific skills involved. Workload has generally been considered to be an important modifier of performance under a variety of personnel and situational conditions, and the assumption that workload can be measured as a unitary concept is central to many decisions affecting the design of man-machine systems. Thus, development of a methodology that yields reliable and valid indices of operator workload should lead to important gains in safety and productivity through resultant modifications of systems designs and operating procedures.

There have been two general approaches to the definition and measurement of workload. One approach has focused on energy expenditure, or effort, as the central concept. This approach has generally attempted to measure workload by using indices based on biomedical measures or on subjective reports of effort. The other approach has focused on the measurement of performance or of system output under the assumption that operators under an increased workload will not be able to perform as well.

One method of studying workload under the latter approach has been through the use of secondary or loading tasks. Knowles (8) summarized early work of this sort and provided the general rationale for the application of the technique to workload measurement. The basic approach in this method is to compare the level of performance achieved on the secondary task, when performed alone, with the level achieved at that task when it is performed in combination with the primary task. Kelley and Wargo (7) noted that, in most tasks, operators will tend to form their own criterion as to how well a given task should be performed and then vary their expenditure of effort up or down to meet it. Once this criterion is established, performance measures on that task become relatively invariant, since insofar as possible, operators will adjust their level of effort to maintain the criterion level of performance. Therefore, the secondary task is added to the primary task in order to increase the subject's workload and to obtain an index of the amount of spare time that the subject has while performing the primary task.

The present study is aimed at developing an objective method for scaling different levels of workload. The Civil Aeromedical Institute's Multiple Task Performance Battery (MTPB) was used to provide several tasks (monitoring lights, monitoring meters, two-dimensional compensatory tracking, pattern identification, mental arithmetic, and problem solving) in different combinations to generate varying job demands and presumably varying levels of workload. Previous studies (2,3,4) with the MTPB have demonstrated that, under various sorts of stress, subjects will tend to protect their performance on the four more-active tasks (i.e., the tasks that require action more often on the part of the subjects; viz, pattern identification, mental arithmetic, problem solving, and tracking), while allowing performance on the monitoring tasks to be negatively affected. Thus, there is a strong tendency for subjects being tested on the MTPB to treat the MTPB tasks as a primary/secondary task combination even though they are instructed to treat all tasks as equally important. In the present study, the monitoring tasks are considered to be secondary tasks, while the four more-active tasks of the MTPB are considered the primary tasks.

II. Method.

Subjects. A total of 94 subjects were tested. All were paid male volunteers between the ages of 18 and 32. The subjects were divided into two groups: Group I (51 subjects) was tested on 2 days, with the same testing schedule for both days, and Group II (43 subjects) was tested on 1 day only.

Performance Battery. The MTPB consists of six tasks that can be presented singly or in any combination. It is computerized so that all signals, problems, and display changes are presented under program control, and all scoring of times and responses are stored for off-line analysis. The same problems or intersignal intervals were presented to all subjects in corresponding testing sessions. Brief descriptions of the nature and performance demands of the tasks follow.

1. Red and green lights monitoring (LT). Pairs of integral lights/switches are located at each corner and in the center of the subject panels. The upper light/switch of each pair is red and the lower is green. Normally, the red lights are off and the green lights are on. A signal on this task consists of a change of state of one of the 10 lights, and a response is made by pushing the appropriate switch; this action returns the light to its normal state. Signals are introduced at random intervals, averaging one signal per minute; response time is recorded separately for the red and green lights. Signals that are not responded to are removed after 15 s, and the response time on that signal is scored as 15 s.

2. Meter monitoring (MTR). The display for this task consists of four edge-reading meters mounted near the top of the subject panel with two pushbutton switches mounted beneath each meter. Normally, the meter pointers are moving at random around an average position of zero (center). When a

signal is introduced, the pointer movement continues as before, except that the average position of the pointer is shifted either to the left or right by an amount that is approximately equal to the maximum excursion resulting from the random movement. The subject responds to a signal on a given meter by pressing the button under that meter on the same side as the signal. When any meter response button is depressed, the random movement is removed and the pointer of that meter stops on its average value, thus giving immediate feedback as to the accuracy of the response. Signals are presented at an average rate of one per minute, and a given signal remains until it is responded to or until it is replaced by the next signal. Response time is computed as the average time the signal is present on the meters; i.e., if a subject does not respond to a given signal, the time that signal is present is included in the response time for the succeeding signal.

3. Mental arithmetic (MATH). The problems for this task are presented on a Burroughs self-scan display mounted at the bottom center of the subject panel. All of the problems are of the form, A + B - C = ?, and are made up of numbers from 11 to 99. The subjects respond by entering their answers on a reverse-order serial entry keyboard; it requires that the least significant digit be entered first. The answer is displayed on the screen as it is entered from the keyboard and may be cleared and changed by the subject. When the subject has entered what he considers to be the correct answer, he depresses the "complete" button. At that time the problem and answer are removed from the screen and the subject is given feedback as to the accuracy of his answer ("R" for right, "W" for wrong). New problems are presented at 20-s intervals. Response time is scored from the introduction of a problem to the time when the subject presses the "complete" button. Accuracy is computed as the proportion of correct answers to total problems presented.

4. Pattern identification (PID). The upper left portion of the Burroughs display is used to present the six-column by six-row patterns in this task. All patterns are in the form of vertical bargraphs with each column height from one through six appearing just once. The problems on this task are analogous to questions on a multiple-choice examination. The first pattern for a given problem is the standard or "question" pattern. This pattern is followed by two comparison patterns presented in succession. The subject must decide if one, neither, or both of the comparison patterns were the same as the standard. Answers are indicated by depressing one of three buttons. On entering an answer, feedback is provided by displaying the correct answer on the screen. The standard pattern appears for 5 s, and each comparison pattern appears for 2 s with 1 s between patterns. New problems are presented every 30 s. Speed of response (from the onset of the second comparison pattern) and accuracy (proportion of correct responses to total problems presented) are recorded.

5. Problem solving (PS). Each subject panel is equipped with five pushbutton switches, a white "task active" light, and three "feedback" lights. The task requires the subject to discover the correct sequence in which to press the five buttons. Each button appears only once in a given solution.

Any time a button is pressed, the amber light is illuminated to show that the response has been acknowledged by the system. A red light provides error feedback and is illuminated whenever the subject makes a response that is not on the correct solution sequence, and turned off when the response is on the correct sequence. Once the subject has pushed all five buttons in the correct order, a blue light is illuminated for 20 s, indicating that the problem has been solved. The next problem is then presented. Each solution is presented twice in succession, and the subject is expected to reenter the previous solution from memory on the second, or confirmation, presentation. Several measures are derived for this task: (a) the speed of solution of the first presentation of a problem; (b) the speed of reentering the solution in the confirmation phase, (c) the proportion of redundant responses made during the solution phase (responses made when information already acquired should make the subject aware that the response being made is not correct), and (d) proportion of error responses made on the confirmation entry of the solution.

6. Tracking (TRK). The display for the tracking task is a cathode-ray tube (CRT) mounted on the upper-center of the subject panel. The target is a dot of light on the CRT, and the center of the CRT is defined by horizontal and vertical crosshairs on the screen. The subject's task is to use a control stick to attempt to counteract a random forcing function and keep the target as near the center of the screen as possible. The forcing function changes the direction of target movement every 3 s. Performance of the tracking task is scored by analog circuitry that integrates absolute error and error squared for each dimension. The error-squared measure is converted to RMS (root-mean-square) error, and vector RMS and vector absolute error measures are derived from horizontal and vertical error scores. Since these measures are highly intercorrelated, vector RMS error is used as a single index of tracking performance.

Procedure. All subjects were trained for 1 hour and then tested in a 2-hour session in which the six MTPB tasks were presented individually for 15 minutes each and the two monitoring tasks were presented together for two 15-minute intervals (LT/MTR-1 and LT/MTR-2). Following 1 or more hours of rest, the subjects were tested for 2½ hours on five complex tasks (30 minutes on each): (a) PS/TRK, (b) MATH/PS, (c) PID/TRK, (d) PID/PS, and (e) MATH/TRK. All five complex-task combinations also included the light-monitoring and meter-monitoring tasks. The same testing schedule was repeated for Group I on the following day (Group II was measured only on the first day). The 1-hour training period was considered sufficient for purposes of this study, but considerably longer periods are generally required for subjects to reach stable performance on the complex tasks (3). Consequently, it was anticipated that scores for Group I would show some improvement on the second day of testing.

III. Results and Discussion.

The differences between the 2 days of Group I's performance were tested with an analysis of variance computed for each of the 12 measures of

TABLE 1. Mean Performance on Combinations of Six Tasks by Two Groups

Measure No.	Task Measure	LT 1	MTR 2	LT/MTR-1 3	LT/MTR-2 4	PS/TRK 5	MATH/PS 6	PID/TRK 7	PS/PID 8	MATH/TRK 9
		Successive Intervals of Performance								
1	Green Lights (LT)	2.21*	-	3.50	4.11	6.43	8.22	5.71	8.10	5.55
	(Response Time)	2.03	-	3.80	3.62	6.58	7.93	6.85	7.77	6.50
2	Red Lights (LT)	1.91	-	2.92	2.77	3.69	4.75	2.71	4.41	3.10
	(Response Time)	1.66	-	2.57	3.38	3.82	4.09	2.96	4.08	2.77
3	Meter Monitoring (MTR)	-	12.78	16.13	15.93	37.94	83.00	56.71	116.90	104.72
	(Response Time)	-	13.38	17.51	17.38	43.38	119.03	85.32	132.48	116.51
4	Mental Arithmetic (Math)	-	-	-	-	-	12.36	-	-	10.15
	(Solution Time)	-	-	-	-	-	11.92	-	-	9.80
5	Mental Arithmetic (Math)	-	-	-	-	-	67.23	-	-	82.11
	(% Accuracy)	-	-	-	-	-	70.94	-	-	82.37
6	Problem Solving (PS)	-	-	-	-	17.86	22.14	-	21.84	-
	(Time/Problem)	-	-	-	-	47.23	22.93	-	22.22	-
7	Problem Solving (PS)	-	-	-	-	88.08	86.52	-	85.17	-
	(% Accuracy)	-	-	-	-	81.11	79.54	-	78.34	-
8	Problem Solving (PS)	-	-	-	-	8.90	13.76	-	11.53	-
	(Confirmation Time/Problem)	-	-	-	-	8.20	15.04	-	10.74	-
9	Problem Solving (PS)	-	-	-	-	82.64	75.16	-	75.22	-
	(Confirmation Accuracy)	-	-	-	-	83.47	72.96	-	78.60	-
10	Pattern Identification (PID)	-	-	-	-	-	-	2.25	2.84	-
	(Response Time)	-	-	-	-	-	-	2.19	2.95	-
11	Pattern Identification (PID)	-	-	-	-	-	-	87.05	76.52	-
	(Accuracy)	-	-	-	-	-	-	87.59	72.59	-
12	Tracking (TRK)	-	-	-	-	108.60	-	90.27	-	98.41
	(RMS Error)	-	-	-	-	102.82	-	85.40	-	97.99

*Upper figure of each pair is mean of Group I, Day 1; lower figure is mean of Group II.

78 11 27 004

performance. Statistically significant ($p < .05$) improvements of Day 2's performances over Day 1 were found, as expected, with several measures; PS speed in all conditions, PS accuracy in the PS/TRK condition, MATH speed and accuracy in the MATH/PS condition, and green lights in the MATH/PS and PID/PS conditions. In no case was Day 2's performance worse than that of Day 1.

The means of the performances obtained on Day 1 with Group I (upper figure in each pair) and with Group II (lower figure) are shown in Table 1 for the nine intervals of performance. The specific task combination employed in any given interval is indicated by the presence of scores for that task in the table.

An analysis of variance was computed with each of the 12 task-performance measures to test for differences between the two groups. Group I was significantly ($p < .05$) better than Group II on PS (accuracy), measure #7 in the table, under all four of the intervals in which that task was included. No other statistically significant difference was found between the two groups.

The reliability of each of the 12 measures was computed with the data of the two groups separately (Day 1 only for Group I). The results are given in Table 2. In each case, the coefficient of correlation for a given measure is the intraclass correlation of that measure from all the task combinations (intervals) in which it appeared. Thus, the intraclass correlation coefficient reported is equal to the mean of all intercorrelations among the measures.

Each of the 24 reliability coefficients given in Table 2 is statistically significant ($p < .05$). With the exception of MTR (response time), measure #3, with Group I where the reliability was .21, the coefficients ranged between .42 (measure #2, Group I) and .91 (measure #4, Group II). These reliabilities are comparable to those found in recent studies with the MTPB (5,6), as well as those found with other versions of the battery (cf. 1, pp. 167-170). In subsequent analyses, the data of the two groups (Day 1 of Group I, and the 1 day of performance of Group II) were combined and treated as a single group.

Each of the 12 measures of performance was tested with analysis of variance for differences across the task combinations (intervals) in which it was employed. The between-intervals variance was statistically significant ($p < .01$) in each of the 12 cases, and it can be inferred that the performance of <u>each</u> task was affected by the combination of tasks with which it was performed.

The data of the monitoring tasks (measures #1, #2, and #3, in Table 1) which appeared in all intervals and which were considered to act as secondary tasks in the complex situations (intervals 5 through 9) were used in a Thurstone Case V scaling procedure (9) to develop a scale of workload for the different task combinations. Data from single-task performance (LT without MTR in interval 1, and the converse in interval 2) and combined monitoring

TABLE 2. Reliability (Intraclass Correlation Coefficients*)

Computed Across MTPB Task Combinations**

Measure	Task (and Measure)	Experimental Group I (Day 1)	Experimental Group II
1	Green LT (Response Time)	.51	.46
2	Red LT (Response Time)	.42	.49
3	MTR (Response Time)	.21	.59
4	Math (Solution Time)	.73	.91
5	Math (Accuracy)	.82	.77
6	PS (Time/Prob)	.59	.69
7	PS (% Accuracy)	.85	.56
8	PS (Confirmation Time/Prob)	.67	.61
9	PS (Confirmation Accuracy)	.80	.52
10	PID (Response Time)	.60	.61
11	PID (Accuracy)	.59	.70
12	TRK (RMS Error)	.64	.73

* All 24 coefficients of correlation are statistically significant ($p < .05$ in each case).

**The different task combinations employed in the 9 performance intervals are shown in Table 1.

performance (intervals 3 and 4) were included with data from the complex-task combinations of intervals 5 through 9 in the scaling procedure. The level of performance was assumed to be <u>inverse</u> to workload; i.e., the greater the workload, the poorer the performance on the secondary, or monitoring, tasks. Thus, the scaling was an inverse scale of performance and a direct scale of workload--the higher the scale value, the lower the performance and the higher the workload represented by the task combination.

Identical scaling procedures were applied to three separate sets of data: Group I's Day 1 and Day 2 data, and Group II's data. The scaling was accomplished by comparing the performance of a given monitoring task (measures #1, #2, and #3) under each task combination (interval), including the single-task performances, with those obtained under all other task combinations (intervals). In each case, the proportion of subjects who performed better under the given condition was noted, and these proportions were then converted to normal-deviate ($\underline{z}$) scores by use of a table of probabilities associated with the normal distribution. The normal-deviate scores were then reflected; i.e., multiplied by -1, and the mean $\underline{z}$ score associated with each task combination was computed. The most negative mean thus represented the best performance and the presumed lowest workload, so within each measure, the largest negative value was subtracted from each of the means, thereby providing a score of zero for the condition with the best performance and lowest workload, and increasing positive scale values for lesser performances and greater workloads. For each of the three monitoring measures, the "zero workload" condition was that of single-task performance. A mean of the three scale scores associated with each interval (task combination) was computed to provide a single workload scale value (WSV) for each interval. The results of the scaling are reported in the WSV rows of Table 3.

TABLE 3. Workload Scale Values (WSV) and

Adjusted Workload Scale Values (AWSV)

		LT/ MTR	PS/ TRK	MATH/ PS	PID/ TRK	PID/ PS	MATH/ TRK
Group I:	WSV	0.52	1.72	2.40	1.61	2.27	1.71
(Day 1)	AWSV	0.00	1.20	1.88	1.09	1.75	1.19
Group I:	WSV	0.49	1.47	1.91	1.31	1.85	1.41
(Day 2)	AWSV	0.00	0.98	1.42	0.82	1.36	0.92
Group II:	WSV	0.56	1.84	2.54	1.90	2.45	2.03
	AWSV	0.00	1.28	1.98	1.34	1.89	1.47

Although single-task performance scores on the LT monitoring and MTR monitoring tasks were each assigned a WSV of zero, this should not be interpreted to mean that such single-task performance is actually a no-workload condition. On the contrary, the zero WSVs assigned to single-task performance represent an arbitrary origin for the scale values. The individual WSVs calculated for the intervals LT/MTR-1 and LT/MTR-2 are independent estimates of the workload imposed by concurrent performance of these two monitoring tasks. The WSVs associated with the five complex task conditions represent concurrent performance of the two active tasks indicated in each case, and also concurrent performance of the two monitoring tasks. Thus, the WSVs associated with the complex performance conditions must be further adjusted if they are to represent only the workload of the condition attributable to the active tasks. Therefore, the mean WSV of LT/MTR-1 and LT/MTR-2 was subtracted from the WSV associated with each of the complex tasks to obtain an adjusted workload scale value (AWSV) for each complex performance condition. This AWSV reflects the workload of combinations of the active tasks after adjustment for the workload imposed by the monitoring tasks. These values are presented in the AWSV rows of Table 3.

The scale values obtained for the five "complex" task combinations were highly correlated among the three sets: (a) .993 for Day 1 versus Day 2 for Group I, (b) .973 for Day 1 Group I versus Group II, and (c) .947 for Day 2 Group I versus Group II. The variabilities of the scores within the three sets of data were essentially identical, but the means differed in agreement with the prior finding that Group I was better on Day 2 than on Day 1, and better on some measures on the first day than Group II. The high WSVs obtained for the combinations MATH/PS and PID/PS (see Table 3) are consistent with the informal reports of subjects and the observations and impressions of the experimenters; subjects who commented on tasks generally mentioned these as being, respectively, the most difficult.

In making a determination of workload, it would be useful in some circumstances if the contribution of individual tasks in a task complex could be evaluated. Such values could be used, for example, to estimate the workload of new combinations of tasks. The simplest model for analyzing the workload attributable to single tasks is to assume additivity; i.e., that particular combinations of tasks do not interact in unique ways and that a given task will impose the same degree of workload regardless of its combination with other tasks. This model is consistent with an analysis of workload in terms of "spare time." To test this model, the AWSVs were analyzed for the workload contribution of the individual active task, using the assumption that the workload imposed by these tasks was linearly additive. Specifically, task workload scale values (TWSVs) were derived by a modification of the method presented by Clark (4) as follows: The AWSVs from each data set were placed in a symmetric array, with rows and columns representing each of the active tasks. Then each row of the array was used to generate a linear equation, assuming that each cell value represents the sum of two unknown variables associated with the tasks presented in that combination. For example, the first row might contain the three AWSVs associated with complex tasks that included tracking. So, the equation associated with that row would be: 3*TRK + MATH + PS + PID = sum of AWSVs in first row. The four simultaneous equations thus generated were then solved to provide the TWSVs presented in Table 4.

TABLE 4. Workload Scale Values Associated With Active Tasks (TWSV)

	TRK	MATH	PID	PS
Group I (Day 1)	0.2625	0.9350	0.8200	0.9375
Group I (Day 2)	0.2300	0.6800	0.6000	0.7500
Group II	0.3750	1.0850	0.9750	0.9050

The method employed to produce the TWSVs in Table 4 provides values that reproduce the row and column values exactly, as well as least squares approximations to the cell values. These derived workload values for individual tasks can be combined by simple addition, under the assumption of linear additivity to provide "predictions" of the workloads imposed by the different combinations of tasks. The "predicted" workload scores derived from the TWSVs are presented in Table 5, along with the AWSVs that they should predict. It may be noted that the predictions closely approximate the original values and thus, in this case, it seems that the workload imposed by the active tasks may be considered to be linearly additive. This is not to say that the assumption of linear additivity is always likely to be appropriate. In fact it seems probable that workload would not be additive for some particular combinations of tasks (cf. 10). In those cases, however, it would be possible to identify the nonadditive combinations by comparing the estimated and observed values and noting where relatively greater divergences occur.

TABLE 5. Actual Adjusted Workload Scale Values (AWSV) vs.

Predicted Values (Based on TWSV Sums)

		PS/TRK	MATH/ PS	PID/ TRK	PID/ PS	MATH/ TRK	RMS Error
Group I	AWSV	1.20	1.88	1.09	1.75	1.19	
(Day 1)	Predicted	1.20	1.87	1.08	1.76	1.20	.0067
Group I	AWSV	0.98	1.42	0.82	1.36	0.92	
(Day 2)	Predicted	0.98	1.43	0.83	1.35	0.91	.0089
Group II	AWSV	1.28	1.98	1.34	1.89	1.47	
	Predicted	1.28	1.99	1.35	1.88	1.46	.0089

The workload attributable to the various active-task combinations was not uniform across the three monitoring tasks (although this is not evident in the summary data presented here). Combinations that involved TRK had the smallest effect on performances of all three monitoring tasks. Combinations that involved MATH had the largest negative effect on MTR, whereas those that involved PS had their largest negative effect on the LT (the sole exception was with green lights, Group II). This suggests that the outcome of the scaling procedure used here and recommended as a method of establishing indices of workload, is dependent not only on the difficulties of the primary tasks that are being scaled, but also to some extent on the nature of the secondary tasks used for the scaling procedure.

Since the contributions of individual tasks to the workload of task combinations could be reliably estimated with the method employed here, it seems safe to infer that similar scaling procedures could be validly applied to predict workloads for task combinations in other studies. Depending on the combinatorial nature of the tasks (e.g., whether they may be considered additive or not), the methodology could be applied to prediction of workload where no test data are available for these combinations; provided, of course, that appropriate data are available for the individual tasks involved (e.g., TWSVs as in Table 4). The method and its potential utility are sufficiently promising to warrant further development and study.

IV. Summary.

A scale of workload was derived for five complex task-combination conditions. The scale provided reliable values that were stable on replication. Where the assumption can be made that the tasks combine in a linearly additive manner, with substantially no task-combination interactions, the scale can be used to estimate both (a) the relative workload contribution of each of the tasks performed in the several task combinations, and (b) the resultant workloads of combinations involving those tasks, including possibly combinations other than those from which the data were derived. The methodology should be applicable to other measures as well, e.g., to biomedical indices of stress or to subjective ratings. The major restriction to the method's use is the requirement that 50 or more subjects be employed in order to yield stable scale values (cf. 9). The availability and use of a valid index of workload would result in gains in both safety and productivity by providing clearer specifications of the demands that are placed on operators under different conditions. The present technique has provided valid indices in this laboratory study. Should it prove to be reliable and valid in operational situations, its use to provide workload specifications should be quite beneficial to the design of both systems and operating procedures.

REFERENCES

1. Chiles, W. D., E. A. Alluisi, and O. S. Adams: Work Schedules and Performance During Confinement. HUMAN FACTORS, 10(2):143-196, 1968.

2. Chiles, W. D., H. L. Gibbons, and P. W. Smith: Effects of Two Common Medications on Complex Performance. FAA Office of Aviation Medicine Report No. 69-9, 1969.

3. Chiles, W. D., and A. E. Jennings: Effects of Alcohol on Complex Performance. HUMAN FACTORS, 12(6):605-612, 1970.

4. Clark, J. A.: A Method of Scaling With Incomplete Pair Comparison Data. EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 37:603-611, 1977.

5. Higgins, E. A., W. D. Chiles, J. M. McKenzie, P. F. Iampietro, J. A. Vaughan, G. E. Funkhouser, M. J. Barr, A. E. Jennings, and G. West: The Effects of Dextroamphetamine on Physiological Responses and Complex Performance During Sleep Loss. FAA Office of Aviation Medicine Report No. 75-14, 1975.

6. Jennings, A. E., and W. D. Chiles: An Investigation of Time-Sharing Ability as a Factor in Complex Performance. HUMAN FACTORS, 19(6):535-547, 1977.

7. Kelley, C. R., and M. J. Wargo: Cross Adaptive Operator Loading Tasks. HUMAN FACTORS, 9(5):395-404, 1967.

8. Knowles, W. B.: Operator Loading Tasks. HUMAN FACTORS, 5:155-161, 1963.

9. Torgerson, W. S.: *Theory and Methods of Scaling*. John Wiley & Sons, Inc., New York, 1958.

10. Wingert, J. W.: Function Interlace Modifications to Analytic Workload Predictions. *In* K. D. Gross and J. D. McGrath (Eds.), Crew System Design, Anacapa Sciences, Inc., Santa Barbara, California, 1973.

☆ U. S. GOVERNMENT PRINTING OFFICE 1978—621-161/169