TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FAA-AM-72-21		

4. Title and Subtitle METHODOLOGY IN THE MEASUREMENT OF COMPLEX HUMAN PERFORMANCE: TWO-DIMENSIONAL COMPENSATORY TRACKING	5. Report Date May 1972 6. Performing Organization Code
7. Author(s) Alan E. Jennings, B.A., W. Dean Chiles, Ph.D., and Georgetta West, B.S.	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. 11. Contract or Grant No.
12. Sponsoring Agency Name and Address Office of Aviation Medicine Federal Aviation Administration 800 Independence Avenue, S.W. Washington, D. C. 20591	13. Type of Report and Period Covered OAM Report 14. Sponsoring Agency Code

15. Supplementary Notes

This research was conducted under Task No. AM-A-72-PSY-13.

16. Abstract

Nineteen subjects were tested on two successive days on a complex performance device designed to measure functions of relevance to aircrew performance; included were measures of monitoring, information processing, pattern discrimination, and group problem solving. The effects of a perceptual-motor-tracking task were evaluated by measuring performance with and without concurrent tracking during five different task combinations and in a tracking-alone condition. The tracking task was shown to be reliable both when performed by itself and when performed concurrently with other tasks. The tracking task was also shown to be sensitive to work load effects from the other tasks and to impose a significant effect on some of the other tasks. The findings also suggest that a composite score based on all concurrently performed tasks may have unique value and sensitivity under some conditions.

17. Key Words	18. Distribution Statement
Complex Performance Time Sharing Tracking Perceptual Motor Skill	Availability is unlimited. Document may be released to the National Technical Information Service, Springfield, Virginia 22151, for sale to the public.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	8	$3.00

Form DOT F 1700.7 (8-69)

METHODOLOGY IN THE MEASUREMENT OF COMPLEX HUMAN PERFORMANCE: TWO-DIMENSIONAL COMPENSATORY TRACKING

I. Introduction.

The Multiple Task Performance Battery (MTPB) was developed for the Air Force as an instrument for testing complex human performance of the sort characterized by aviation operations (Chiles, Alluisi, and Adams, 1968). The MTPB was originally used to measure the influence of work-rest cycles on performance, and it has since been used in evaluating the effect on performance of such factors as acute alcohol intoxication (Chiles and Jennings, 1972), common medications (Chiles, Gibbons, and Smith, 1969), infectious diseases (Alluisi, Thurmond, and Coates, 1971), and other variables. MTPB performance has also been shown to be significantly correlated with instructor evaluations of the potential of Air Traffic Controller trainees to become fully rated controllers (Chiles, Jennings, and West, 1972).

The basic strategy of the MTPB is to require the time-shared performance of a number of relatively simple tasks presented in varying combinations. These tasks, which were designed to assess aviation-related performance skills, are described in detail in a later section of this report, but, for present purposes, they may be grouped into two categories: monitoring tasks and active tasks. The monitoring tasks require the subject to scan a number of lights and meters and respond to a non-normal state of a particular light or meter; typically, these tasks have been operated continuously during MTPB testing, either alone or as a background for the active tasks.

Two of the three active tasks, mental arithmetic and visual pattern identification, are timed with new problems presented at fixed time intervals. The third is a group problem solving task that requires trial-and-error solution and is subject-paced except for a fixed interval between a correct solution and the introduction of the next problem.

All these tasks share a common factor in that they require only intermittent attention and fairly simple, discrete responses. On all tasks except arithmetic, the response is accomplished by pushing a single button or throwing a lever switch, while for the arithmetic task a three-digit answer is entered into a keyboard and a lever switch is thrown. However, many real-world operator situations, such as flying an airplane or driving an automobile, require essentially continuous, graded control over some function (e.g., controlling attitude or steering) as well as monitoring (e.g., monitoring engine instruments or the brake lights of a preceding car) and problem solving (e.g., estimating ground speed or distance required to pass). This consideration indicated the desirability of adding some sort of perceptual-motor control task to the MTPB in order to assess a broader range of operator functions. Recently this addition was made in the form of a two-dimensional compensatory tracking task, which is the focus of the present study.

The tracking task was used as a part of the MTPB in a recent unpublished study (conducted by Education and Public Affairs, Inc., for the FAA under Contract DOT–FA–70WA–2371) in which the MTPB, along with a number of other tests, was evaluated as a possible selection device for Air Traffic Controllers. In this study, 258 Air Traffic Controllers were tested on the MTPB for one hour each. The tracking task and the monitoring task were presented during the final 15 minutes of testing. Throughout the study, signals were presented on the monitoring tasks at a somewhat higher rate than in the present study. The final block of 10 minutes of tracking performance was divided into two 5-minute subperiods. The product moment cor-

Assistance rendered by Ralph Guild and Dale Renke is gratefully acknowledged.

relation between the scores for these two subperiods was computed for each of the tracking measures. The reliabilities thus obtained varied between a high of .73 to a low of .68 with Spearman-Brown predicted 10-minute reliabilities ranging from .84 to .81.

Past experience has shown that the reliability and sensitivity of MTPB scores were influenced by the particular task constellation presented. Also, it had been determined that some task combinations were unreasonable in that the subjects were overloaded to the point that their performance became erratic. Therefore, the present study was undertaken to provide information on the reliability of the tracking task when performed as a part of various task constellations and to examine the effect of tracking on the other tasks in terms of both reliability and mean scores. This information is necessary to the efficient design of experiments in which the tracking task is to be used. It will provide guidelines in the selection of task combinations and test session durations.

In earlier studies, the scores from each task were analyzed individually, and the conclusions drawn were based on the pattern of results obtained. More recently, in the work with Air Traffic Service trainees, it was found that a composite score, consisting of a variance-equalized mean of individual task scores, gave good validities against an instructor evaluation criterion. Therefore, the analysis in the present study will consider both individual task scores and composite performance scores.

II. Method.

A. *Subjects.* The 19 subjects who served in this experiment were all male college students in their twenties. None had previous experience with the test apparatus. Each subject was paid a total of $20 for participating in the experiment.

B. *Apparatus.* The subjects were tested on the Multiple Task Performance Battery (MTPB). The MTPB is designed to test subjects on a number of simple tasks which may be presented in isolation or concurrently in any combination. The tasks in the MTPB include the newly installed two-dimensional compensatory tracking task; two monitoring tasks—warning lights and meter monitoring; a mental arithmetic task; a group problem solving task; and a pattern identification task. Up to five subjects may be tested simultaneously on the MTPB. Each subject is seated in a booth that contains an intercom and an MTPB subject panel on which are located the displays and response switches for the various tasks.

C. *Warning Lights.* The warning lights task consists of two subtasks, red and green lights. These lights are mounted in five pairs, one pair in each corner and one in the center of the subject panels. The red lights were normally off; whenever one of the red lights was illuminated the subject was to turn the light off by pressing a button immediately beneath the light. Similarly, the green lights were normally on, and, when one went out, the subject was to re-illuminate it by pressing the button immediately beneath the light. Response times were measured separately for the red and green lights from the onset (or offset) of a light until the subject returned the light to its normal state. If the subject failed to respond to one of the warning lights, the light would return to its normal state after 30 seconds and the subject was given a response time of 30 seconds for that signal. The warning light signals were introduced at random time intervals with a mean intersignal interval of one minute.

D. *Meters.* The meter task required the subject to monitor four meters mounted in a row across the top of his panel. Normally, the meters fluctuated around a mean position of 0 (12 o'clock position). When a signal was introduced, the mean position of a given meter would shift 25 units (about 15°) either to the left or right. This change in the mean position was approximately equal to the maximum excursion of the meter's random motion. When the subject detected a shift in the average reading on a meter, he responded by throwing a three-position, spring-centered, lever switch, located beneath the meter, in the direction of the bias. A bias would stay on a meter until the subject responded correctly to it or until a new signal was introduced. Response time on this task was taken from the introduction of a signal until the subject responded correctly, either to that signal or to some succeeding signal. For example, assume that a signal was introduced and the subject did not respond to it; assume that after 70 seconds that signal was replaced by a new signal.

If the subject responded to the new signal in 20 seconds, he would be given a detection time of 90 seconds for that response. Meter signals were introduced at random time intervals with a mean intersignal interval of one minute.

E. *Arithmetic.* In the mental arithmetic task, three numbers were displayed across the bottom of the subject's panel; the subject was required to compute the sum of the first two numbers and subtract the third number from that sum. Subjects entered their answer on a set of three decade push buttons mounted on top of their panels; they indicated that they had completed the problem by throwing a lever switch. Problems were presented at a fixed rate of three per minute. Feedback information was provided by a blue light mounted next to the arithmetic display; the light flashed at the end of each problem if the subject had responded correctly. Performance was measured in terms of mean response time and the percentage of correct responses.

F. *Pattern Identification.* The display for the pattern identification task is a screen located in the lower left corner of the subject's panel. Behind the screen is a six by six matrix of lights, part of which is illuminated to form a pattern. All patterns were in the form of a vertical bar graph with six bars varying between one and six units in height. For each problem, three patterns are presented in succession; a standard pattern was presented for five seconds, and two comparison patterns were presented for two seconds each. The subject was to decide whether the first, the second, or neither comparison pattern was the same as the standard. He indicated his answer by pressing one of three buttons marked "1," "2," or "N" (Neither). Feedback information was provided by three blue lights mounted above the response buttons; at the end of each problem the light over the correct response button would flash. Problems were presented at a rate of two per minute. Performance was measured in terms of percentage of correct responses.

G. *Group Problem Solving.* The group problem solving task required the subjects to discover the correct sequence in which each subject should push a button located on his panel. A correct solution required each subject to push his button once and only once in the correct sequence. A red light located on each subject's panel provided feedback information by being turned on whenever an incorrect response was made. A standard trial-and-error search procedure was used to discover the solution sequence. When a problem had been successfully solved, a green light was illuminated for 22 seconds; then the next problem was started. Each problem was presented on two successive trials; thus the task had two phases; in the first phase, the problem was solved by utilizing the search sequence, and, in the second phase, the previously discovered solution was reentered from memory. Response time on this task was based on the immediately preceding problem solving event, that is, from either the introduction of a new problem or from the button-push that preceded the response. Time-per-response was computed separately for first and second solution, and, also, the percentage of correct responses during the second solution phase was computed.

H. *Tracking.* The display for the tracking task consisted of a 7–cm cathode ray tube as an integral part of an oscilloscope. The scope was mounted directly above the subject's panel at about eye level. Vertical and horizontal cross-hairs defined a zero-error position at the center of the screen. The controlled element appeared on the screen as a 1–mm dot of greenish-white light. All controls for the scope were masked from the subjects except for the position controls, which were adjusted by the subject before each testing session to correct for eye height and position. The control for this task consisted of a small joy stick mounted so it could be easily controlled with the right hand. The stick was adjusted to give zero output when it was vertical. A 20° deflection of the stick in either dimension resulted in a 3–cm movement of the dot; forward movement of the stick moved the dot down; movement to the right moved the dot to the right. While the subjects were tracking, a forcing function was introduced which moved the dot randomly about the screen. The subjects were to counteract this disturbance by manipulating the joy stick to keep the dot in the center of the screen. The maximum excursion of the dot due to the forcing function was 3.5 cm in either dimension; the dot could not go beyond the "reach" of control. The forcing function and, to a lesser extent, the control had a damping circuit added so that the dot moved to its new position in accordance with an "S"-shaped time-position curve. The tracking task was controlled and scored by means of analog circuitry which provided independent scoring for the horizontal and vertical dimensions for each subject. The forcing function was generated as a step function by choosing randomly among positions on a voltage divider every three seconds. This voltage was then put through a "shaping circuit" which, in effect, converted the series of square waves to a complex sine wave. The forcing function and stick output voltages were summed algebraically and the resultant value defined the position of the dot on the scope face and informed the subject as to his momentary tracking error. The momentary error voltage was converted to an absolute value and integrated to yield the integrated absolute error. The absolute error was also run through a squaring circuit and integrated to yield an error-squared measure that was used to calculate RMS error. Both types of error measure were allowed to accumulate for one minute, at the end of which time the values were printed out and the integrators reset to zero. The subject's score on tracking was the mean integrated error per minute for the final 10 minutes of tracking in a given 15-minute scoring interval. In all, six measures of tracking performance were used: horizontal and vertical absolute error; horizontal and vertical RMS error (the square root of the mean error-squared scores); the vector sum of the absolute error, computed by squaring the two absolute error scores, summing them and taking the square root of that sum; and the RMS vector sum, computed by simply adding the two integrated error-squared values and taking the square root of their sum.

I. *Composite.* A composite measure based on all but the tracking scores was computed for each task combination by converting the raw scores to a standard score format and taking the mean standard score for all measures from a given interval. The conversion to standard scores was accomplished by a linear transformation to a metric with a mean of five and a standard deviation of one; better performance was indicated by scores above the mean.

The inclusion of the active tasks in this composite precluded its use in testing differences between task combinations. Therefore, a second composite, based on the three monitoring tasks, was computed in the same way for use in between task combination comparisons.

III. Procedure.

The subjects were tested in four groups of five subjects each on two successive afternoons. In the second group, only four subjects were tested and one of the experimenters acted as a "stand-in" for the fifth subject in the group problem solving task. Before the first test session the subjects received approximately one hour of orientation and training on the MTPB tasks. The purpose of the experiment was first explained to the subjects. The three monitoring tasks were then demonstrated and set to introduce signals at the rate to be used during testing; this was done to give the subjects practice in scanning the panel for signals. The other tasks were then introduced and explained one at a time; the subjects were allowed sufficient practice on a given task to assure that they understood what was required before proceeding to the next task. The subjects were presented the following amount of practice: 10 two-digit arithmetic problems, 10 pattern identification problems, eight group problem solving problems (first and second solution equal one problem), and about two minutes of tracking practice. At the end of the training period, the experimenter answered any questions and outlined the schedule of task combinations that would be presented in that session. The subjects were then reminded that insofar as possible, all tasks should be given equal priority, and the test was started. During testing, the subjects were observed from an adjoining room through a half-silvered mirror.

In all, 11 different task combinations were presented in this study. One of these consisted of tracking presented in isolation. The other 10 consisted of five different combinations of the other tasks, each presented with and without concurrent tracking. The five task combinations were: monitoring alone; pattern identification with monitoring; group problem solving with monitoring; and arithmetic at two difficulty levels, each with monitoring. The easier arithmetic problems required manipulation of one digit elements while the more difficult arithmetic problems consisted of two-digit elements similar to the problems presented during training.

The five task combinations were presented in a counter-balanced order that was different for every testing session. The presentation order was arranged so that the two arithmetic conditions were not presented successively. Each task combination was presented for 30 minutes, 15 minutes with concurrent tracking and 15 minutes without. Two groups had tracking during the first 15 minutes and two during the second 15 minutes. This order was retained across both sessions for a given group.

The temporal placement of the tracking-alone condition in the testing schedule was such that the subjects were never required to track for longer than 15 minutes at a time.

The entire testing schedule required 2 hours and 45 minutes per session. The session was extended to 3 hours by the addition of a 15-minute break after 1 hour 30 minutes or 1 hour 15 minutes, depending on whether the tracking-alone condition was scheduled in the first or second half of the testing session.

During the time interval when the subjects were not tracking, the control sticks were disabled and a constant voltage was applied to the forcing function input sufficient to drive the dot off-screen. When the subjects were to begin tracking, the dot would first move to the center of the screen; after 10 seconds, the control sticks were activated, and, after another 10 seconds, the forcing function was introduced and scoring started. If any subjects did not begin tracking, they were alerted via intercom.

At the beginning of the tracking-only condition, the subjects were alerted via intercom and the random movement of the probability meters was turned off. The beginning of the other task combinations was indicated fairly unambiguously by the activity of the task displays.

IV. Results.

The data, both raw scores and composites, were analyzed to obtain reliability estimates for each task. Differences between task conditions were examined (where applicable), as were differences within task conditions (tracking and practice effects).

An overall index of task reliability was obtained for each individual measure and for the composite score by correlating the sum of all day-one scores for a given measure with the sum of all day-two scores for that measure. In addition, day-one vs. day-two test-retest reliabilities were computed for each measure for the 15 minutes of performance of a given task combination.

The reliabilities for the tracking measures are reported in Table I. In general, the overall reliability was about as good as that found for the other tasks and, with the exception of the monitoring conditions, was generally constant across conditions. The monitoring condition resulted in markedly lower reliabilities for all tracking measures; this was due, at least in part, to a single subject who did very poorly in the second session on tracking with monitoring. This condition fell in the final 15 minutes of the testing session for that group, and this subject was apparently not attending to the tasks during this interval. However, even when this subject

TABLE I.—Tracking Task Test-Retest Reliability by Task Condition

	Sum	Tracking	Monitoring	One-digit Arithmetic	Two-digit Arithmetic	Group Problem Solving	Pattern Identification
Horizontal Absolute	.72*	.47	.08	.66*	.60*	.46	.75*
Vertical Absolute	.75*	.58*	.43	.75*	.55*	.64*	.55*
Absolute Vector	.76*	.58*	.32	.78*	.54*	.54*	.68*
Horizontal RMS	.68*	.45	.15	.65*	.66*	.46	.81*
Vertical RMS	.67*	.57*	.55*	.60*	.55*	.59*	.60*
RMS Vector	.74*	.55*	.44	.64*	.58*	.56*	.71*

* $p < .05$

was dropped from the analysis, the tracking reliabilities were still lower for the monitoring condition than for the other conditions.

The monitoring and composite reliabilities are summarized in Table II. In addition to the overall sum reliability, separate reliabilities were also computed for each measure for the sums of all performance with concurrent tracking and without concurrent tracking. Reliabilities were very high for red lights, good for meters, but rather low for the green lights. The reliability of the sum of the with-tracking scores tended to be lower than that of the without-tracking sums on the green lights, meters, and the composites, but in no case did the reliabilities for the two conditions differ significantly. During the two-digit arithmetic condition both meter detection times and the overall composite score were significantly more reliable without than with tracking. None of the other conditions displayed a consistent effect of the with- and without-tracking condition.

In the one-digit arithmetic condition, the meters task was significantly more reliable with than without concurrent tracking. In the pattern identification condition, the meters task was significantly more reliable without tracking but the red lights task was more reliable with tracking. In the monitoring condition, red lights were significantly more reliable without tracking.

Table II.—Monitoring Task and Composite Test-Retest Reliability by Task and Tracking Condition

		Sum		Monitoring		One-digit Arithmetic		Two-digit Arithmetic		Group Problem Solving		Pattern Identi-fication	
	Sum	With	WO	With	WO	With	WO	With	WO	With	WO	With	WO
Red Lights	.98*	.97*	.95*	.54*	.85*	.78*	.90*	.89*	.82*	.75*	.75*	.86*	.44
Green Lights	.36	.16	.51*	.31	.34	.36	.09	-.08	.26	.10	-.09	.15	.41
Meters	.70*	.64*	.83*	.89*	.75*	.74*	.37	.39	.73*	.53*	.60*	.54*	.86*
Monitoring Comp	.68*	.60*	.82*	.86*	.68*	.79*	.58*	.39	.58*	.30	.33	.77*	.50
Composite	.87*	.78*	.91*	.86*	.68*	.80*	.67*	.27	.78*	.28	.52*	.81*	.58*

* p<.05 members of underlined pairs of r's differ at p<.05

Table III reports the reliabilities of the active tasks. All were fairly good except for two of the group problem solving task measures, namely, second solution time-per-response and percentage correct. The reliability of the second solution time measure was probably depressed because this measure is based on relatively few responses. Typically, only six to eight problems were completed in a 15-minute interval, and, unless there was an error, each subject made only one second solution response per problem. The second solution percentage-correct measure, besides being based on an equally small number of responses, suffered from a ceiling effect in that relatively few errors were made on the second solutions; thus, for most of the scoring intervals, most of the subjects made 100% correct responses.

Table III.—Active Tasks Test-Retest Reliability

	Sum	With Tracking	Without Tracking
One-digit Arithmetic Time	.73*	.71*	.67*
percent	.71*	.65*	.63*
Two-digit Arithmetic time	.80*	.48	.86*
percent	.59*	.50	.44
Pattern Id. percent	.67*	.67*	.50
Group Task first solution time	.64*	.73*	.42
second solution time	.46	.08	.62
second solution percent	−.01	.04	.09

* p<.05
members of underlined pairs differ at p<.05

A. *Between Condition Differences.* The tracking scores, monitoring scores, and monitoring composite scores were tested for differences between task combinations using a two-step approach. First, an analysis of variance model was applied with task condition included as a source of variance. Where a significant task combination effect was found, differences between all pairs of means were tested by Tukey's HSD test.

The analysis of variance of the tracking measures indicated a task combination effect significant at $p<.001$ for each measure. The analysis of variance for the monitoring composite also showed a significant effect of task combinations ($p<.001$). The task effect in separate analyses of the three individual monitoring measures was significant at the .01 level for each measure. The results of the pair-wise comparisons are summarized in Table IV.

The pattern of significant differences was found to be similar for all six tracking measures, and, therefore, the tracking measures are represented by a single matrix in Table IV.

On all measures, tracking performance was best during the tracking-only condition, followed in order by monitoring, group problem solving, pattern identification, one-digit arithmetic, and two-digit arithmetic. Also for tracking performance, the tracking-only, monitoring, and two-digit arithmetic conditions were significantly different from the other three conditions and from each other. Tracking performance did not differ across the group problem solving, pattern identification, and one-digit arithmetic conditions. The single exception to this general finding was for the RMS vector sum; for this measure there was no difference between the one-digit and two-digit arithmetic conditions.

TABLE IV.—Pairwise Comparison of Tracking and Monitoring Scores Between Task Conditions

Tracking Measures	M	G	P	A1	A2
Tracking Condition (T)	*	*	*	*	*
Monitoring Condition (M)		*	*	*	*
Group Problem Solving Condition (G)			–	–	*
Pattern Identification Condition (P)				–	*
One-digit Arithmetic Condition (A1)					*
Two-digit Arithmetic Condition (A2)					

Red Lights Response Time		P	A1	G	A2
	M	–	–	*	*
	P		–	–	–
	A1			–	–
	G				–
	A2				

Green Lights Response Time		P	A2	A1	G
	M	*	*	*	*
	P		–	–	*
	A2			–	–
	A1				–
	G				

Meters Response Time		A2	A1	P	G
	M	*	*	*	*
	A2		–	–	*
	A1			–	*
	P				*
	G				

Monitoring Composite		P	A1	A2	G
	M	*	*	*	*
	P		–	–	*
	A1			–	*
	A2				*
	G				

* $p<.05$ In all tables, conditions are ranked from best (top)
– $p>.05$ to worst (bottom)

It should be pointed out that the intercorrelations among the tracking measures are quite high. Correlations between absolute error and RMS error from the same condition ranged from a high of .98 to a low of .87 with a mean correlation of .96. Correlations between absolute vector and RMS vector were somewhat higher than this. Correlations between horizontal and vertical mean absolute error scores were between .60 and .88 with a mean of .80 while between horizontal and vertical RMS error scores the range was from .63 to .87 with a mean of .79.

B. *Within Condition Effects.* The tracking measures were analyzed for practice effects in a "condition by practice by subjects" analysis of variance. Neither the main effect of practice nor the interaction term was significant for any of the tracking measures. Inspection of the minute-by-minute raw scores showed that tracking error tended to level off very rapidly during the first "with tracking" interval; typically, most of this "leveling off" took place in the first five minutes of the first interval.

Practice and tracking effects were evaluated separately for each task condition using a "days by tracking (with and without) by subjects" analysis of variance. The mean scores associated with the main effects of these analyses are summarized in Table V. All significant differences were in the expected direction; better performance was found in the second testing session and/or the without-tracking condition. No significant practice or tracking effects were found for the individual scores or the composite for the monitoring-only condition. Comparisons between task combinations showed this condition to be associated with the best monitoring performance; apparently, the monitoring-only condition was easy enough to be relatively insensitive to practice and tracking effects. Both practice and tracking had significant effects on all of the other composite scores.

TABLE V.—Mean Task Scores by Task Condition, Practice, and Tracking Condition

	Composite	Red Lights	Green Lights	Meters			
Monitoring Condition							
Day 1—Day 2	4.92–5.06	2.10–2.28	3.14–2.78	14.57–23.75			
With—WO Tracking	4.99–5.00	2.12–2.27	2.78–3.18	20.92– 9.50			
					Arithmetic		
One-digit Arithmetic Cond.					Time		Percent
Day 1—Day 2	4.81–5.71**	2.79–2.47	5.80–5.05	20.16–20.30	6.54–5.67**		91–93*
With—WO Tracking	4.84–5.14**	2.53–2.78	6.14–4.72*	25.27–15.19	6.44–5.77**		91–93
Two-digit Arithmetic Cond.							
Day 1—Day 2	4.76–5.20**	3.29–2.78*	5.99–4.61*	22.37–17.25	8.31–7.48**		89–91
With—WO Tracking	4.78–5.18**	3.70–2.37*	6.07–4.54*	25.41–14.22	8.22–7.56**		88–93*
					Pattern		
Pattern Id. Condition					Percent		
Day 1—Day 2	4.83–5.15*	3.15–2.43	7.75–5.24	26.23–50.02	93–94		
With—WO Tracking	4.90–5.08*	2.92–2.64	7.15–5.84	58.27–18.28*	92–95*		
					First Solution	Second Solution	
Group Task Condition					Time	Time	Percent
Day 1—Day 2	4.88–5.10*	2.75–2.29*	5.16–4.37*	25.07–20.30	1.95–1.71	1.53–1.36	95–98
With—WO Tracking	4.84–5.14*	2.63–2.40	4.85–4.65	30.88–14.48	1.77–1.89*	1.51–1.38	95–98

*—pairs different at $p < .05$
**—pairs different at $p < .01$

In the one-digit arithmetic condition, both arithmetic measures—time and per cent correct—showed significant practice effects. Tracking effects were demonstrated on green lights, meters, and arithmetic time. The two-digit arithmetic condition yielded similar results, and, in addition, there was a practice effect on red and green lights and a tracking effect on red lights.

In the pattern identification condition, none of the individual tasks showed a significant practice effect, though the composite score did. Meters and per cent correct were significantly affected by tracking, as was the composite. The interaction effect found on green lights was examined further by taking the simple effect of tracking at both practice levels and vice versa. The only significant simple effect was the effect of practice in the with-tracking condition.

Red and green lights showed a significant practice effect during the group problem solving task. For group problem solving, first solution time-per-response was the only individual measure to show a significant tracking effect.

V. Discussion.

The tracking task was shown to be about as reliable as the other MTPB tasks; except for the low reliability coefficients found during the monitoring-only condition, the reliabilities of this task were quite good. The high tracking reliabilities during the simpler tracking-only condition and the relatively high reliabilities of the monitoring tasks during the monitoring-with-tracking condition preclude any adequate explanation of this finding.

The absence of practice effects on tracking makes it appear that performance on this tracking task reaches asymptotic levels or at least a fairly stable plateau very rapidly. However, it should be noted that this is a relatively easy task as compared to most laboratory tracking tasks. In addition, this finding may be an artifact of the design, since the tracking task was clearly central to the study and the subjects may, therefore, have given it a high priority. Thus, the learning effect might be more pronounced if different emphasis were given in the instructions.

The extremely high intercorrelations of the tracking measures, especially between absolute and RMS error, suggest that it is superfluous to use both absolute and RMS measures. The high correlations between vertical and horizontal error also suggest that, for most purposes, the vector sum alone would be an adequate single measure of tracking performance.

The between-condition differences found for the tracking measures seem to be related in a fairly straightforward manner to the amount of time that the subject is required to look away from the tracking task display. That is, during the tracking-only condition, with essentially no distraction, tracking performance was at its best. Adding the monitoring tasks required the subject to periodically scan his panel, but the self-paced nature of this scanning and the fact that relatively little time was required to respond to the monitoring task signals should make for minimal disruption of tracking performance. Although no significant differences were found for tracking performance among the problem solving, pattern identification, and one-digit arithmetic conditions, the rank order of these conditions is consistent with the above hypothesis. The group problem solving task should require the least visual attention since the task is coordinated verbally, and the only time a subject need look at his panel is immediately after he pushes his button. The pattern task demands more visual attention, but, even including the off-time between patterns, the display is active for only 11 out of every 30 seconds. The arithmetic task clearly involves the greatest degree of visual distraction since the display is the most complex and problems are presented at a rate of three per minute.

The fact that the tracking task is sensitive to work load effects suggests the possibility that this task may be fairly sensitive to variables external to the test situation as well. In a recent study by Chiles, Iampietro, and Higgins (1971), a tracking task with the same dynamics as this one was found to be more sensitive than other measures to the effects of altitude and elevated temperature.

Finally, there is no apparent problem in combining the tracking task with the monitoring tasks or, on the whole, with the active tasks. At the training level employed in this study, the two-digit arithmetic condition may be approaching a maximum usable difficulty level. Specifically, the significantly lower composite reliability during two-digit arithmetic with tracking can be interpreted to indicate somewhat more erratic overall performance. However, the tracking reliability remained high during this condition, which suggests that the subjects may have tended to protect tracking performance at the expense of the other tasks.

Comparison of the levels of performance on a given task across different task combinations revealed clearly the dependence of monitoring task performance on the demands of the concurrently performed tasks. For example, the two warning lights tasks showed a significant practice effect in the two-digit arithmetic and group problem solving task conditions. This is presumably not due to an improvement in reaction time *per se*; for warning lights performance during the monitoring conditions, not only was there no significant practice effect, but red lights performance was slightly (though not significantly) better in the first testing session. These practice effects must then be attributed to changes in the performance of other tasks and/or in the time-sharing skill domain.

The absence of a significant practice effect for the meters and the problem solving tasks may have resulted from the way priorities were assigned to the tasks by the subjects. Or, it may have been that the total training and testing times were not sufficient to permit any learning effects to be revealed in the form of performance improvements. This latter possibility is compatible with previous observations of subjects performing these tasks; subjects may not show improvements in performance until after several hours of practice. However, we cannot offer a clear-cut explanation for this finding.

The generally greater reliability and sensitivity of the composite score is most clearly demonstrated in the pattern identification condition. Here, no single score showed a significant practice effect but the composite does. Although the composite cannot be used to replace analysis of individual task scores, it does provide a useful supplement. The primary value of the composite should be in the case where the subjects do not attach the same priorities to the different tasks. It has been consistently shown in complex performance situations that, when the operator is stressed, some function or functions will be maintained at or near pre-stress levels at the expense of other functions (e.g., Chiles and Jennings, 1970). This seems to hold whether the stress is low skill level, work load, or some stressor external to the test situation. The arithmetic task is an example of this; up to a point, arithmetic accuracy is consistently protected at the expense of solution time. However, when subjects hold differing priorities or when all functions are affected about equally but only slightly, the composite would be expected to come into its own. As a hypothetical example, suppose that a given condition affects the performance of all subjects, but each subject's performance is affected only on a single measure and the subjects are about equally divided as to which is the affected measure. In this hypothetical example, the estimate of the strength of effect would be quite low for each individual measure considered by itself, and the error terms for the statistical analyses would be relatively large. On the other hand, the composite score, which would tend to be more stable, would include the affected measure for each subject and, therefore, should be more sensitive to possible stress effects. We interpret the data to suggest that this is what happened with the task combination involving the pattern identification condition where individual measures showed no practice effect but the composite did.

VI. Summary and Conclusions.

This study appears to demonstrate the potential usefulness of the tracking task as a part of the Multiple Task Performance Battery. The tracking task was shown to be reliable both when performed by itself and when performed concurrently with other tasks. In most cases, the addition of the tracking task does not have an adverse effect on the reliability of the concurrently performed tasks. The task also shows promise of being fairly sensitive to stressor conditions, if we are justified in extrapolating from work load stress to stressors external to the task characteristics. It is concluded that the addition of two-dimensional compensatory tracking to the battery will provide performance measures of greater relevance to aviation operations; indeed, the measures should prove useful in assessing performance in most any situation in which the operator is required to exercise the time-shared performance of manual-control along with other types of tasks.

The findings also suggest that a composite score based on all of the MTPB measures provides an overall measure of performance that may have unique value in experimental situations in which different subjects may assign different priorities to the performance of the constituent tasks of the battery.

REFERENCES

1. Alluisi, E. A., J. B. Thurmond, and G. D. Coates: Behavioral Effects of Infectious Diseases: Respiratory Pasteurella Tularensis in Man. PERCEPTUAL AND MOTOR SKILLS, (Monograph Supplement 2–V32) 32:647–668, 1971.
2. Chiles, W. Dean, Harry L. Gibbons, and Paul W. Smith: Effects of Two Common Medications on Complex Performance. FAA Office of Aviation Medicine Report No. AM–69–9, 1969.
3. Chiles, W. Dean, P. F. Iampietro, E. A. Higgins, J. A. Vaughan, Georgetta West, and G. E. Funkhouser: Combined Effects of Altitude and High Temperature on Complex Performance. FAA Office of Aviation Medicine Report No. AM–71–17, 1971.
4. Chiles, W. Dean, and Alan E. Jennings: Effects of Alcohol on Complex Performance. HUMAN FACTORS, 12:605–612, 1970.
5. Chiles, W. Dean, and Alan E. Jennings: Effects of Alcohol on a Problem Solving Task. FAA Office of Aviation Medicine Report No. AM–72–11, 1972.
6. Chiles, W. Dean, Alan E. Jennings, and Georgetta West: Multiple Task Performance as a Predictor of the Potential of Air Traffic Controller Trainees. FAA Office of Aviation Medicine Report No. AM–72–5, 1972.
7. Chiles, W. Dean, E. A. Alluisi, and Oscar S. Adams: Work Schedules and Performance During Confinement. HUMAN FACTORS, 10:143–196, 1968.

23816