FAA-AM-80-17

AN EXPLORATORY INVESTIGATION OF VARIOUS ASSESSMENT INSTRUMENTS AS CORRELATES OF COMPLEX VISUAL MONITORING PERFORMANCE

Richard I. Thackray
R. Mark Touchstone
Civil Aeromedical Institute
Federal Aviation Administration
Oklahoma City, Oklahoma

October 1980

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

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for its content or use thereof.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FAA-AM-80-17		

4. Title and Subtitle	5. Report Date
AN EXPLORATORY INVESTIGATION OF VARIOUS ASSESSMENT INSTRUMENTS AS CORRELATES OF COMPLEX VISUAL MONITORING PERFORMANCE	OCTOBER 1980
	6. Performing Organization Code
7. Author(s) Richard I. Thackray and R. Mark Touchstone	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.
12. Sponsoring Agency Name and Address Office of Aviation Medicine Federal Aviation Administration 800 Independence Avenue, SW. Washington, D.C. 20591	13. Type of Report and Period Covered 14. Sponsoring Agency Code

15. Supplementary Notes

Work was performed under tasks AM-C-79-PSY-77 and AM-C-80-PSY-77.

16. Abstract

The present study examined a variety of possible predictors of complex monitoring performance. The criterion task was designed to resemble that of a highly automated air traffic control radar system containing computer-generated alphanumeric displays. Forty-five men and women were administered a battery of tests and tasks prior to performing the criterion task. Extreme groups, separated on the basis of their performance decrement scores, differed significantly on 6 of the 28 predictor variables. In general, the significant relationships obtained were in accordance with expectations. All correlations were low which agrees with the findings of previous studies of predictors of performance on simple vigilance tasks. Since the criterion task simulated the task requirements of advanced, highly automated air traffic control systems still in the planning stage, the utility of any of the significant predictors in predicting performance of controllers on contemporary systems would require further research using actual performance of present-day controllers on such systems as the criterion.

17. Key Words	18. Distribution Statement
Air Traffic Control Attention Automation Monitoring Performance Selection Vigilance	Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	17	

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

AN EXPLORATORY INVESTIGATION OF VARIOUS ASSESSMENT INSTRUMENTS AS CORRELATES OF COMPLEX VISUAL MONITORING PERFORMANCE

I. Introduction.

Air traffic controllers are required to maintain high levels of alertness and attentiveness to radar displays for relatively long periods of time. While increasing automation will undoubtedly change the nature of the tasks to be monitored, as long as a requirement exists for the controller to back up system operation in the event of a malfunction or emergency, the necessity of maintaining sustained alertness will continue to be demanded if safe operation is to be insured. To the extent that highly automated systems require only infrequent operator intervention, the problem of maintaining high levels of vigilance under such conditions is an important and formidable one indeed.

One of the commonest findings in research on vigilance is that large and relatively consistent differences exist among individuals in their ability to sustain attention over long periods of time (13). An early study by Mackworth (23), for example, found that subjects who detected 92 percent of the signals during the first hour on a simple vigilance task averaged 87 percent during the next 3 hours. Subjects who detected only 70 percent initially, detected less than 50 percent during the remainder of the experiment. This was one of the first studies to demonstrate that stable individual differences in the ability to sustain attention exist, and it pointed to the possibility of classifying individuals with regard to this variable. Subsequent studies by Jenkins (21) and Buckner, Harabedian, and McGrath (4) have found that reliable individual differences exist not only within monitoring sessions but across sessions as well. A disturbing finding in the Buckner et al. (4) study, however, was that consistency of individual performances was not maintained across different <u>types</u> of vigilance tasks, and this result led those investigators to question the existence of a general "vigilance" factor.

If individual differences in vigilance performance are task specific, then the findings of any study concerned with correlates of vigilance performance would have limited applicability to other monitoring tasks that differ from the task studied. Fortunately, a number of more recent investigations have found relatively high correlations between monitoring performance on different tasks when these tasks are comparable in type and in difficulty of signal discrimination (9,17,19,25,33).

Thus, accumulating evidence supports the view that individual differences in monitoring efficiency are both reliable and are maintained across tasks having similar characteristics. Relatively little research has been conducted, however, to determine the pattern of subject characteristics that may contribute to these individual differences in monitoring efficiency. Most of the studies on correlates of vigilance performance have examined only a single subject variable. Other than physiological indices of arousal, the most frequently employed variables have included intelligence, extraversion, age, and sex (see 13,22,27, for reviews). As Davis and Tune (13) note, there is little conclusive evidence that sex, intelligence, or age are significant variables in the performance of simple vigilance tasks. Extraversion, on the other hand, has been found by a number of investigators (1,7, 12,30) to be related to performance on repetitive or vigilance-type tasks, with performance decrement confined largely to extraverted subjects.

In the literature we found only two studies that have explored a wide range of behavioral measures as possible predictors of simple vigilance performance. In the earliest study (24), 54 male subjects performed visual and auditory monitoring tasks on separate occasions over a total of 32 1-hour sessions. The test battery administered to all subjects included the Navy Classification Battery, a variety of aptitude tests, the Guilford-Zimmerman Temperament Survey (16), the Taylor Manifest Anxiety Scale (28), and several tests of motivation. In general, tests which emphasized clerical speed and accuracy and tests of motivational persistence (e.g., the O-dotting test) showed the highest correlations with visual vigilance performance. However, only the O-dotting test correlated significantly with performance during a subsequent cross-validation study.

A more recent study by Waag (34) examined a wide range of measures, including several of those employed by McGrath et al. (24), as possible predictors of performance on a simple visual monitoring task. The sample size was unusually large (478 subjects) in order to achieve the reliabilities required for multivariate analysis. Out of the 67 predictor measures examined, 19 showed low but significant correlations with number of correct detections on the vigilance task. Using a stepwise regression analysis, 12 of these predictor variables contributed significantly to the explanation of the criterion variance with a multiple R of .40. The variables or tests entered into the regression equation included sex of the subject, two scales (16PF A and 16PF C) from the Sixteen Personality Factor Questionnaire (5), the Cognitive Structure and Order scales of the Personality Research Form (20), the MAT C1-C10 and MAT II scales from the Motivation Analysis Test (6), a level of aspiration measure, two measures derived from the O-dotting test, a measure of time estimation, and a measure of speed of visual scanning. On the basis of these predictors, a set of descriptors characterising those individuals who perform well on simple visual vigilance tasks include (a) female sex, (b) somewhat quiet and reserved, (c) emotionally stable, (d) likes order and organization, (e) dislikes uncertainty, (f) has high levels of aspiration, (g) persistent worker, and (h) rapid visual scanner. It should be noted, however, that each of these variables taken singly correlated quite low with the criterion. (The highest correlation was obtained for sex difference, and it accounted for only 4 percent of the criterion variance.)

While the studies by McGrath et al. (24) and Waag (34) represent the most comprehensive attempts to date to examine correlates of vigilance performance, several other studies should also be mentioned. Halcomb and Kirk (18) administered the six scales (Dominance, Achievement via Independence, Social Presence, Responsibility, Flexibility, and Self-Control) of the California Psychological Inventory (15) and an intelligence test, the Wonderlic Personnel Test (36), to subjects who subsequently performed a simple visual monitoring task. Superior vigilance performance was found for subjects scoring high on flexibility or self-control. In addition, those who scored high on the combined achievement and intelligence scales were superior performers.

Thackray, Bailey, and Touchstone (31) explored the general pattern of physiological and subjective changes associated with performance decrement on a complex visual monitoring task. Blood pressure, oral temperature, palmar skin conductance, gross body movement, heart rate, and heart rate variability were measured along with self-ratings of boredom, monotony, irritation, attentiveness, fatigue, and tension.

Performance decrement was associated with high levels of reported boredom and monotony, low attentiveness, high levels of reported tension, and increases in heart rate variability. This study apparently represents the most comprehensive evaluation to date of possible subject correlates of complex monitoring performance. Even so, only physiological variables and subjective experiences were examined.

The present study was basically an extension of the one (31) just described. Its purpose was to examine relationships between a wide variety of possible predictor variables and performance efficiency using a task which simulated the general display characteristics and visual monitoring requirements of the highly automated, advanced air traffic control systems currently being contemplated. Most of the tests and tasks used as predictors were selected from the more promising ones used by McGrath et al. (24), Thackray et al. (31), and Waag (34). Thus, this study sought to (i) determine whether certain measures found to predict monitoring performance on simple vigilance tasks would also predict performance on a complex monitoring task requiring constant scanning and greater information processing, (ii) replicate the findings of our earlier study (31), and (iii) examine additional measures which we believed might be related to complex monitoring performance.

II. Method.

A. Subjects. Forty-five paid volunteers, 26 men and 19 women, served as subjects (Ss). All were obtained from the general population (e.g., college students, housewives) and ranged in age from 18 to 29 years. None had prior experience with the tasks used or previous training in air traffic control. All had normal visual acuity (corrected to 20/20 if necessary), and none had any reported hearing loss.

B. Criterion Task. Task programing and recording of responses were accomplished using a Digital Equipment Corporation (DEC) PDP-11/40 computer. The computer was interfaced with a VT-11 (DEC) 17-inch (43 cm) cathode-ray tube (CRT), which served as the S's display. The CRT was located in a console resembling an air traffic control radar unit. The stimuli (targets) consisted of small rectangular "blips" representing the locations of given aircraft. Adjacent to each target was an alphanumeric data block. Data blocks comprised two rows of symbols: the top row, consisting of two letters and three numerals, identified the aircraft, while the bottom row of six numerals indicated its altitude and speed. The first three of these numerals gave altitude in hundreds of feet and the last three gave groundspeed.

A simulated radar sweepline was employed that made one complete clockwise revolution every 6 seconds. A target was updated as to (i) location and (ii) any change in its data block moments after the sweepline passed the target's prior location. Targets normally moved in a linear fashion unless a course change was necessary to avoid target overlaps. The critical stimulus or signal to which the S was instructed to respond consisted of a change in a target's displayed altitude to a value greater than 550 or less than 150. The values of the increases or decreases in altitude were randomly determined, except that the changed altitude value could not be greater than 599 or less than 100. Ten such critical stimuli appeared in each 30-minute period; five occurred in the first 15 minutes and five in the second. The S's response to a critical stimulus consisted of pressing a button held in the right hand and then holding a light pen over the critical target. The light pen caused the altitude portion of the data block to revert to its previous value. If the S failed to detect a critical stimulus within 1 minute, the data block automatically reverted to its previous value. Marker channels on a Beckman Dynograph signaled the onset of a critical stimulus and the occurrence of the required button press. All performance data were recorded by the computer for subsequent processing.

The same target display file was used for all Ss and was initially constructed from a computer program which assigned an altitude, groundspeed, identification, entry point, and exit point to each of the targets. All assignments were randomly determined except for the following restrictions: (i) altitudes had to fall within the "normal" range of 150 to 550 (in hundreds of feet), (ii) groundspeeds had to fall within the range of 400 to 550 knots, and (iii) entry and exit points of a given target could not be separated by less than 30° along the circumference of the simulated radar screen. In addition, time of critical stimulus occurrence and the target in which it occurred were randomly determined with the restriction that two targets could not contain critical stimuli at the same time.

Performance data were computer processed and the following measures were obtained on each S for each 30-minute period (all latency measures refer to the time from critical stimulus onset to the button press):

(i) Mean response latency to critical stimuli correctly identified.
(ii) Single longest latency to a correctly identified critical stimulus.
(iii) Single shortest latency to a correctly identified critical stimulus.
(iv) Number of critical stimuli missed.

C. Predictors and/or Task Correlates. The following tests, tasks, or variables yielding 28 separate measures were employed as possible correlates of performance on the complex monitoring task:

1. O-Dotting Task (ODOT).--This task was used both by McGrath et al. (24) and by Waag (34). It appears to measure an individual's motivation to persist at what is essentially a meaningless task. The S was instructed simply to place a dot as rapidly as possible in the center of each O on a sheet filled with rows and columns of O's. Subjects worked continuously for 4 minutes, marking their places when a buzzer sounded at the end of each minute. Following a 40-second rest, the S performed for another minute.

From the resulting data, two measures were obtained: (a) performance decrement (PD); that is the mean number of O's dotted during the last 2 minutes minus the mean number of O's dotted during the first 2 minutes, and (b) performance recovery (PR); that is the number of O's dotted during the fifth minute minus the mean of the third and fourth minutes.

2. Target Identification Task (TAR).--This was a slightly modified version of another of the tasks used by Waag (34). Subjects viewed a series of 40 slides each of which consisted of a matrix of 112 letters (14 rows and 8 columns). Thirty-two of the slides contained only T's; the remaining eight contained an I at a random location in place of one of the T's. Each slide was projected for 10 seconds onto a screen located two meters from the S. The subjects were instructed to scan each

slide for a target (I) and press a hand-held button if one was detected. The two performance measures were: (a) the number of targets not detected, i.e., omissions (OM), and (b) the number of errors of commission (CM).

3. Serial Reaction Task (SR).--Several of our previous studies have shown a positive relationship between performance decrement (a progressive increase in response variability) on this task and both self-rated distractibility (29) and extraversion (30). Since this task requires continuous, sustained attention, it appeared desirable to include it in our battery. The S was instructed to press one of four keys in response to a number (1-4) appearing in a visual display. Correct responses caused a different number to appear in the display. This stimulus-response cycle continued as long as correct responses were made; incorrect responses would not initiate a new stimulus until the correct key was pressed. Elapsed time between each successive correct response was electronically measured and the data punched on paper tape. The task was performed continuously for 40 minutes. Two performance measures were obtained: (a) change in variability of response time (ΔSD) from the first to the last 4 minutes of the session, and (b) mean response time ($\bar{X}$RT) across the session.

4. Auditory Habituation (AH).--The electrodermal response to tones habituates with repeated tone presentations. The rate of this habituation has been shown to be related to performance decrement in simple vigilance tasks (8,11,26). In general, individuals who habituate rapidly show greater decrement than do those who habituate slowly. We incorporated electrodermal response habituation in the present study using a 1000 Hz, 70 dB tone presented 20 times at 20- to 40-second intervals. Electrodermal response was measured from electrodes attached to the palmar surfaces of the index and middle fingers of the left hand and recorded as conductance. Subjects were instructed to sit quietly and listen to the tone presented over headphones. The habituation measure was the trial number at which three successive tone presentations failed to evoke a skin conductance response of at least 0.1 micromhos within 4 seconds of tone onset.

5. Time Estimation (TIME).--In the study by Waag (34), estimates of task duration showed some correlation with vigilance decrement. Poor performance tended to be associated with longer estimated task time. In the present study, Ss were told prior to the start of the criterion task that the session length varied somewhat from person to person, but that in no event would it last less than 1½ hours or more than 2½ hours. Immediately following the end of the session, Ss indicated their estimate of task duration by placing a checkmark along a rating scale anchored by 1½ hours at one end and 2½ hours at the other. Actual task duration was always 2 hours. An estimated time score was obtained by measuring the distance of the checkmark along the scale.

6. Eysenck Personality Inventory (EPI).--The Eysenck Personality Inventory (14) is a short inventory that measures introversion-extraversion and neuroticism and has been used in several studies of vigilance performance. Although the findings are not entirely consistent, introverts generally appear better able to sustain attention during prolonged monitoring than do extraverts (1,7,10,12,30). Because of these findings, it was decided to include this inventory in the test battery. Both measures of extraversion (E) and neuroticism (N) were used.

7. Personality Research Form (PRF).--This is a relatively new test devised by Jackson (20) to measure personality traits relevant to the functioning of individuals in a variety of situations. Five of the 20 scales were selected for inclusion in the battery of predictors. These scales were Achievement (AC) (a measure of striving, accomplishing, purposeful, attaining behavior), Endurance (EN) (willingness to persevere and persist in work habits), Cognitive Structure (CS) (tendency to be precise, meticulous, accurate, and perfectionistic), Change (CH) (a measure of the desire to avoid routine, unchanging circumstances), and Impulsivity (IM) (hasty, rash, uninhibited, impatient tendencies). The first three of these scales were also used by Waag (34) and found to correlate significantly with simple vigilance performance.

8. Sensation Seeking Scale (SSS).--The SSS (37) was originally developed to measure individual differences in the need for varied sensory stimulation. Unpublished results of several of our previous studies suggested slight but nonsignificant, relationships between perceived boredom--monotony during repetitive stimulation and scores on the SSS. Consequently, we decided to explore further the relationship of SSS scores to task performance. Three measures derived from this scale were included in the test battery: total score on the SSS scale (GS) and scores on two subscales (Boredom Susceptibility (BS) and Experience Seeking (ES)).

9. Subjective Rating Scale (SRS).--The SRS is a simple self-rating scale that we developed and have used in several previous studies. It consists of six 9-point scales measuring the dimensions of attentiveness (ATT), fatigue (FAT), tension (TEN), boredom (BOR), irritation (IRR), and monotony (MON). The SRS was studied most extensively in the experiment by Thackray, Bailey, and Touchstone (31). In that study, Ss falling at the extremes of rated boredom and monotony following exposure to a simulated radar monitoring task were compared with respect to several physiological, performance, and subjective variables. In general, those who rated the task as quite boring and monotonous showed the greatest decline in rated attentiveness, the largest decrement in performance, and the greatest increase in heart rate variability. Because the SRS was highly related to indices of declining attention, we included the SRS in the present battery.

10. Heart Rate Measures (HR).--Measures of change in mean heart rate and in heart rate variability have been routinely included in several of our previous studies (29,30,31,32), with change in heart rate variability frequently showing a relationship to vigilance decrement (30,31). Measures of cardiac activity were incorporated in the present study by taking the difference between each S's mean heart rate during the first and last half hour of criterion task performance ($\Delta\bar{X}$), as well as the difference in heart rate variability (ΔSD) during the same periods. In addition, mean heart rate and mean heart rate variability were obtained across the entire task session.

D. Procedure. On arrival the S was taken to one of the two experimental rooms and administered the various predictor tests or tasks in the following order:

1. Serial Reaction Task
2. Eysenck Personality Inventory
3. Target Identification Task
4. Sensation Seeking Scale
5. Auditory Habituation
6. O-Dotting Task
7. Personality Research Form

All Ss were administered these tests or tasks in the same order during the morning with a 5-minute break each hour. Total testing time was approximately 3 hours.

After lunch, the S was taken to the room containing the simulated radar (criterion) task and instrumented for heart rate recording. The 9-point subjective rating scale was then administered dealing with present feelings of attentiveness, fatigue, tension, irritation, and boredom.

The task instructions emphasized the necessity of pressing the button immediately upon detection of a critical stimulus. The S was told that a critical stimulus (any altitude value greater than 550 or less than 150) could occur in any target at any time, regardless of the current altitude values of the targets. It was explained that occasional large changes in altitude would not normally occur in an actual radar system, but that this departure from normal conditions was necessary to insure that all targets would be given equal priority in scanning. Following the taped instructions, the S was given a 4-minute practice period containing six critical stimuli.

After the 2-hour task session, the S estimated the duration of the task period and completed a second form of the subjective rating scale. This form was identical to the first except that the S was asked to rate each item, plus one additional item dealing with task monotony, on the basis of how the S felt near the end of the task period just completed.

III. Results.

A. Criterion Measures. Figure 1 shows mean detection latencies across 30-minute periods for all critical stimuli, as well as mean maximum and minimum latencies, for the total group of 45 Ss. Also shown are the standard deviations for each measure. The general trends for maximum, minimum, and mean latencies are essentially the same as those obtained in several of our previous studies under comparable experimental conditions. Performance is relatively uniform during the first hour, but becomes progressively worse during the second. Analyses of variance applied to these three sets of data revealed significant main effects for the four 30-minute periods for mean latencies, $F(3/132) = 10.40$, $p < .01$ and for maximum latencies, $F(3/132) = 2.82$, $p < .05$. The change in minimum latencies was not significant, $F(3/132) = 2.00$, $p > .05$.

Although significant changes were obtained for both maximum and mean latencies, estimates of the reliabilities of these two measures based upon analyses of variance (35) yielded quite different values for maximum and mean latencies (.29 and .76, respectively). Consequently, of the two measures, only mean latencies were used as criterion measures in subsequent analyses.

Errors of omission were another possible criterion variable. However, as shown in Table 1, most Ss either missed no stimuli or, at the most, missed only one. Because of the nature of this distribution, it did not appear feasible to employ errors of omission as a criterion measure.

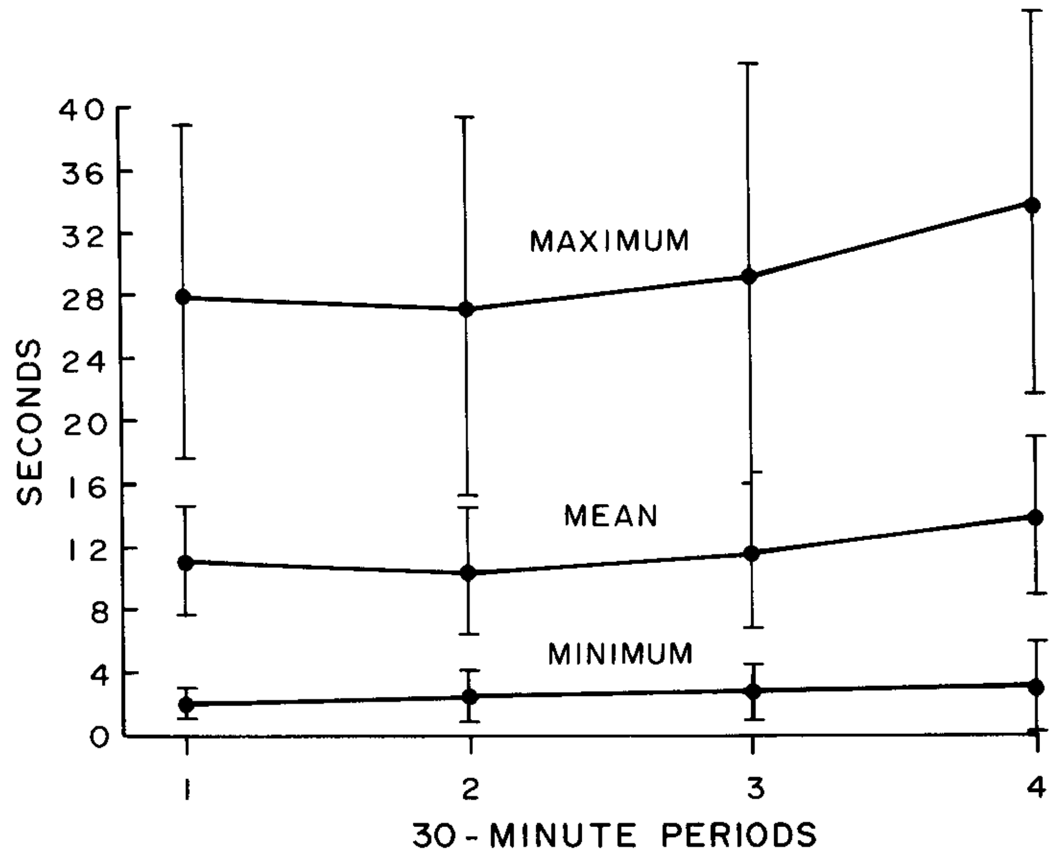

FIGURE 1. Mean, maximum, and minimum detection latencies and their standard deviations for the total group of Ss.

TABLE 1. Frequency Distribution for Number of Critical Stimuli Missed

Number	Frequency	Percentage
0	22	49
1	13	29
2	3	7
3	3	7
4	3	7
5	0	0
6	0	0
7	0	0
8	1	2

B. Establishment of Criterion Groups. Each S's mean detection latency during the first half hour was subtracted from his/her mean latency during the last half hour. From the resulting distribution two extreme groups of 10 Ss each were formed. Figure 2 shows mean detection latencies for the two groups across the 2-hour session.

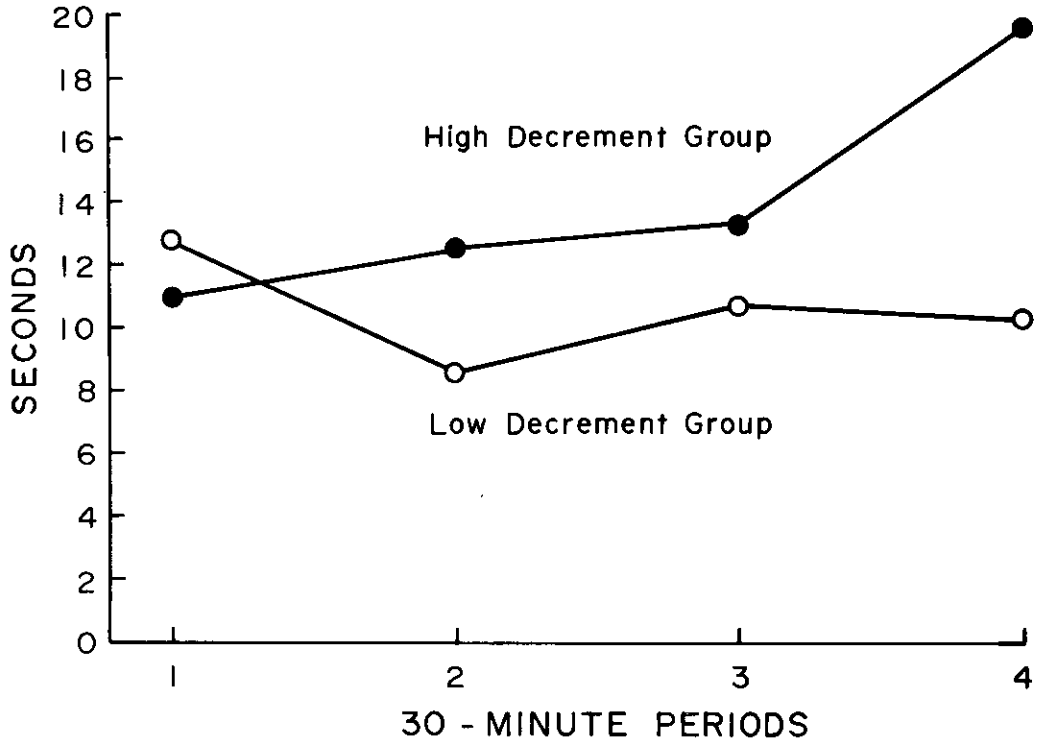

FIGURE 2. Mean detection latencies for the high and low decrement groups.

Mean increase for the high latency group was 8.1 seconds, while the low latency group showed a mean decrease of -2.6 seconds. An analysis of variance of these data revealed a significant main effect for 30-minute periods, $F(3/54) = 7.13$, $p < .01$ and a significant interaction effect, $F(3/54) = 12.44$, $p < .01$. The main effect for groups was not significant, $F(1/18) = 3.20$, $p > .05$.

A second dichotomous classification was obtained by taking each S's mean detection latency across the 2-hour session and then selecting the 10 Ss with the longest overall latencies and the 10 Ss with the shortest latencies. Mean detection latencies were 16.5 and 8.1 seconds for the slow and fast group, respectively.

These two methods of classifying Ss were not entirely independent, as shown by the correlation of .36 ($p < .05$) between mean latency scores and decrement scores for the total group of Ss. However, in a correlation of this size, 87 percent of the variance in overall speed of detection would be independent of decrement scores. Consequently, the classification based upon mean latencies was retained in the hopes that it might reveal somewhat different relationships with the predictor measures than those shown by classification according to decrement scores alone.

C. Relationships of Predictor Variables to Criterion Groups. t-tests were performed on each of the predictor variables according to the two *a priori* methods of classifying the criterion groups. The means, standard deviations, and t-values for groups separated on the basis of decrement scores and overall detection latencies are shown in Tables 2 and 3, respectively. Because of the exploratory nature of this study, t-values which had associated probability values of 10 percent or less were considered to be significant.

TABLE 2. Summary of t-tests for High and Low Decrement Groups

Predictor Variable	Mean of High Group	SD of High Group	Mean of Low Group	SD of Low Group	t Value
ODOT PD	9.30	4.98	2.90	12.20	1.46
ODOT PR	24.80	8.10	18.85	12.36	1.21
TAR OM	3.00	1.48	1.70	1.49	1.86*
TAR CM	1.10	1.51	1.00	1.18	0.16
SR ΔSD	0.06	0.05	-0.01	0.09	1.92*
SR $\bar{X}$RT	0.77	0.15	0.81	0.14	0.66
AH	3.20	3.31	4.30	3.38	0.70
TIME	70.20	36.01	51.00	33.73	1.17
EPI E	13.80	4.71	14.30	4.90	0.22
EPI N	13.00	4.96	14.50	4.48	0.67
PRF AC	10.80	3.31	9.70	2.33	0.82
PRF EN	10.20	2.14	11.20	2.23	0.97
PRF CS	9.10	3.11	9.20	3.25	0.07
PRF CH	9.80	2.89	10.70	2.37	0.72
PRF IM	9.00	3.82	6.00	2.90	1.88*
SSS GS	12.00	2.90	11.40	3.26	0.41
SSS BS	7.10	1.87	6.40	1.80	0.81
SSS ES	7.60	2.20	9.90	2.02	2.31**
SRS ATT	4.20	1.45	4.00	1.41	0.29
SRS FAT	6.70	1.27	7.00	1.10	0.54
SRS TEN	2.70	1.42	3.30	2.00	0.73
SRS BOR	5.80	2.23	4.80	2.40	0.92
SRS IRR	1.40	0.80	2.40	1.62	1.66
SRS MON	6.20	1.72	4.30	1.85	2.26**
HR $\Delta\bar{X}$	-3.57	2.99	-6.45	4.65	1.56
HRV ΔSD	0.66	0.83	1.47	0.69	2.24**
$\bar{X}$HR	77.75	6.93	85.18	12.33	1.58
$\bar{X}$HRV	7.73	1.66	7.20	1.54	0.71

*p < .10

**p < .05

TABLE 3. Summary of t-tests for High and Low Groups Separated According to Overall Mean Detection Latencies

Predictor Variable	Mean of High Group	SD of High Group	Mean of Low Group	SD of Low Group	t Value
ODOT PD	6.45	6.99	7.20	8.15	0.21
ODOT PR	20.00	9.62	21.30	10.52	0.27
TAR OM	3.40	1.28	2.10	1.45	2.02*
TAR CM	2.10	2.02	0.70	0.90	1.90*
SR $\underline{\Delta}$SD	0.07	0.10	0.06	0.08	0.39
SR $\bar{X}$RT	0.82	0.16	0.72	0.06	1.71
AH	5.50	3.80	3.60	3.29	1.13
TIME	54.80	33.17	69.80	34.55	0.94
EPI E	13.20	4.87	11.40	4.22	0.84
EPI N	12.50	5.28	12.30	4.50	0.09
PRF AC	9.50	2.84	10.10	2.62	0.47
PRF EN	9.90	1.92	10.60	2.24	0.71
PRF CS	9.10	3.01	8.10	3.51	0.65
PRF CH	9.20	3.76	9.60	2.15	0.28
PRF IM	7.10	3.42	7.30	3.82	0.12
SSS GS	13.10	2.62	11.40	3.10	1.25
SSS BS	6.50	2.84	6.70	2.10	0.17
SSS ES	8.50	3.04	8.00	3.77	0.31
SRS ATT	4.10	1.22	3.90	1.64	0.29
SRS FAT	6.50	0.81	6.90	1.22	0.82
SRS TEN	3.20	1.89	3.90	1.37	0.90
SRS BOR	5.10	2.12	4.70	2.28	0.39
SRS IRR	1.30	0.64	2.20	1.78	1.43
SRS MON	4.90	1.81	5.40	2.46	0.49
HR $\Delta\bar{X}$	-4.88	3.36	-2.75	2.69	1.48
HRV ΔSD	0.99	1.50	1.76	0.92	1.32
$\bar{X}$HR	80.28	10.50	76.62	9.63	0.77
$\bar{X}$HRV	6.75	1.02	76.62	7.63	1.43

*$\underline{p} < .10$

As is evident in Table 2, those $\underline{S}$s showing the greatest and those showing the least performance decrement differed significantly on 6 of the 28 predictor variables. The HRV ΔSD scores require some explanation, since the greater change in HRV for the low than for the high decrement group would appear to be contrary to previous findings (30,31). Figure 3 plots HRV for the two groups across the 2-hour session. An analysis of variance of these data revealed a nonsignificant main effect for groups, $F < 1.00$, but a significant effect for 30-minute periods, $F(3/54) = 17.41$, $\underline{p} < .01$ and a significant interaction effect, $F(3/54) = 2.58$, $\underline{p} < .05$. Thus, while HRV increased for both groups, it was significantly depressed in the low relative to the high decrement group for most of the 2-hour session.

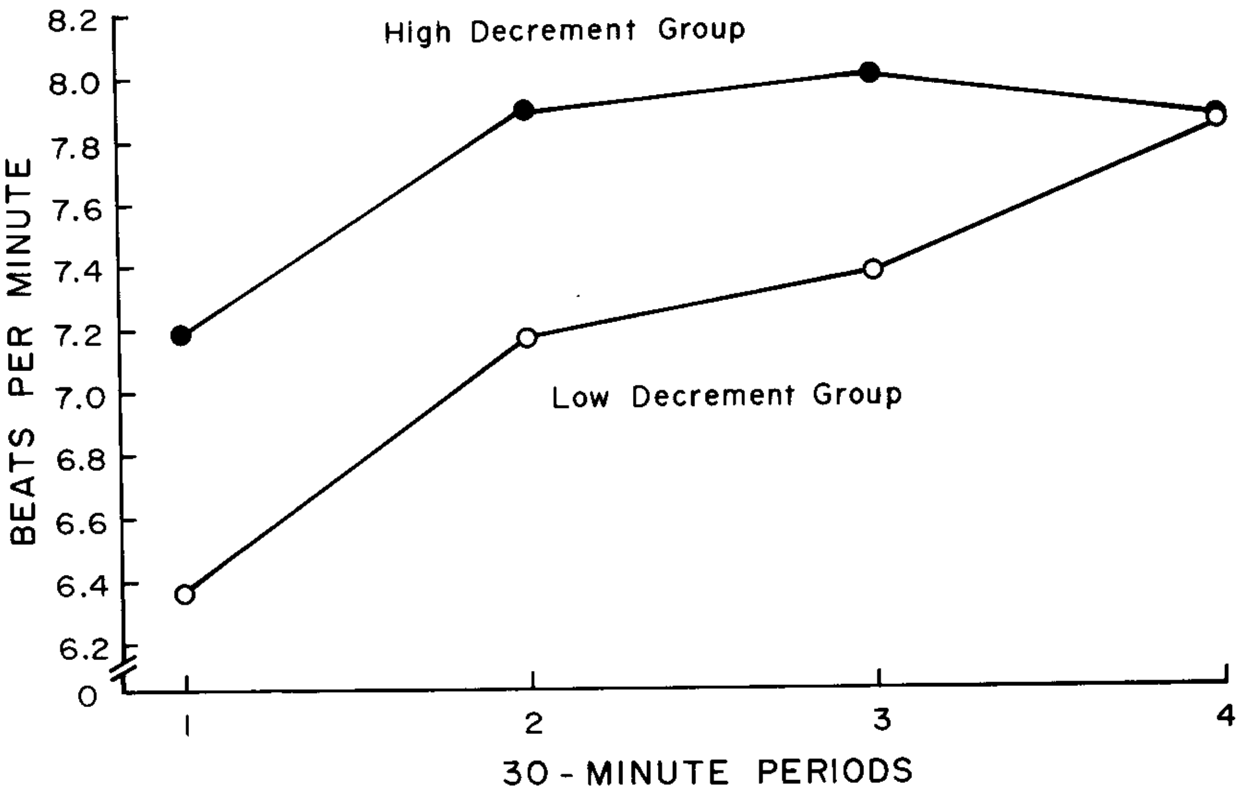

FIGURE 3. Heart rate variability for the high and low decrement groups.

Subjects classified according to overall mean detection latency differed on only two of the predictor variables. These data are shown in Table 3. One additional variable, SR $\overline{XRT}$, approached, but did not reach significance at the 10 percent level. It is apparent from a comparison of Tables 2 and 3 that the predictor variables employed were more closely related to performance decrement than to speed of stimulus detection per se.

For each of the significant relationships shown in Tables 2 and 3, product moment correlations were computed between predictor and criterion scores using the total group of Ss. Correlations of TAR OM, SR ΔSD, PRF IM, SSS ES, SRS MON, and HRV ΔSD with decrement scores yielded correlations of 0.27, 0.14, 0.23, -0.20, 0.17, and -0.27, respectively. Only TAR OM and HRV ΔSD were significantly related to the criterion measure ($p < .10$). In Table 3, both TAR OM and TAR CM were significantly correlated with mean latencies ($r = 0.36$, $p < .05$ and $r = 0.42$, $p < .01$, respectively).

IV. Discussion.

Extreme groups of Ss, formed on the basis of their change in detection latencies during 2-hour performance of the criterion task, were found to differ significantly on six of the predictor variables. Those Ss showing the greatest performance decrement on the radar monitoring task, as contrasted with those Ss showing minimal decrement or even improved performance, were found to (a) show greater decrement on a repetitive, perceptual-motor task (the Serial Reaction Task), (b) omit more targets in the Target Identification Task, (c) score higher on the Impulsivity Scale of the Personality Research Form, (d) rate the criterion task as being more monotonous, (e) show greater heart rate variability during performance of the criterion task, and (f) score higher on Experience Seeking on the Sensation Seeking Scale.

Three of the above relationships were in general accordance with the findings of previous studies conducted in our laboratory. Thus, heart rate variability, self-rated task monotony, and impulsivity have all been shown to be related to performance decrement on simple repetitive and/or complex monitoring tasks (29,30, 31).

The relationship between increased performance variability on the Serial Reaction Task and performance decrement on the criterion task confirms one of our previous assumptions. Increased variability in serial reaction performance appears to be a manifestation of an increase in lapses of attention (2,3,22), and our earlier research using this task was predicated on the assumption that such lapses should be related to measures reflecting variability of attention on more complex monitoring tasks. The apparent relationship between these two different types of tasks, both requiring sustained attention, suggests the existence of a common attentional-type factor.

Further support for an attentional factor is suggested by the finding that Ss falling at the extremes in performance decrement differed significantly in the number of targets missed in the Target Identification Task. Errors of omission on a similar identification task were also found by Waag (34) to discriminate between good and poor performers on a simple vigilance task. In fact, Waag found this to be one of the best predictors of vigilance performance. Considering the nature of the Target Identification Task, this is not too surprising. Effective performance on this task requires sustained attention and rapid scanning. The first of these is a requirement for low error rates on simple vigilance tasks, while both are required for effective performance on a more complex monitoring task.

The finding that Experience Seeking on the Sensation Seeking Scale was related to performance decrement is difficult to explain. If the significant finding indicates a real relationship, it would suggest that those individuals who are better able to sustain attention to a monitoring task are more likely to seek out novel, unconventional, and exciting experiences than those who find it difficult to sustain task attention. Since unpublished data from one of our earlier studies (31) found only a slight, nonsignificant relationship between monitoring performance and scores on this particular scale, and in a direction opposite to that found in the present study, it seems likely that the relationship obtained in the present study was fortuitous.

The various predictors were relatively insensitive in differentiating between groups separated on the basis of their mean detection latencies. Significant differences were obtained for only two variables: errors of omission and errors of commission, both of which were derived from the Target Identification Task. It will be recalled that errors of omission also differentiated between groups separated according to decrement scores.

Although six predictors were significantly related to the high and low decrement groups and two to the extreme latency groups, correlations of these predictors with either decrement or mean latency scores of the total group of Ss were disappointingly low. Only errors of omission on the Target Identification Task correlated significantly with both decrement scores ($r = .27$, $p < .10$) and mean latency scores ($r = .36$, $p < .10$). As in the study by Waag (34), neither of these correlations is large enough to be used singly for prediction in any practical situation. However, as noted previously, target identification was also one of the best predictors found in the study by Waag (34). This suggests that some version of this predictor, possibly a paper and pencil one, might be usefully combined with existing predictors in air traffic controller applications.

V. Conclusions.

The present study was exploratory in nature and was designed to examine further some of the relationships to complex monitoring performance found in our previous studies, as well as to explore other possible correlates of this type of performance. In general, the significant relationships obtained were in accordance with our expectations. Most of these relationships were low which agrees with the findings of Waag (34). Even the best predictor found in this study, errors of omission on the Target Identification Task, accounted, at best, for only about 13 percent of the variance in performance on the criterion task. Consequently, none of the significant predictors would be useful by themselves as predictors of complex monitoring performance.

No attempt was made to use multiple regression procedures to develop a prediction equation. This would have been inappropriate in view of the number of predictors or correlates examined in relation to the sample size employed. It should be recognized also that the criterion task was designed to simulate the task characteristics of a level of automation which does not yet exist in present air traffic control systems. The simulated task called for sustained attention, relatively minimal information processing, and continuous scanning activity. These are obviously only some of the skills or abilities required by present controllers. While such predictors as the Target Identification Task might prove useful (when combined with other predictors of known value) in predicting performance of controllers on contemporary systems, this could only be determined through further research using actual performance of present-day controllers on contemporary systems as the criterion.

REFERENCES

1. Bakan, P.: Extraversion-Introversion and Improvement in an Auditory Vigilance Task. BRITISH JOURNAL OF PSYCHOLOGY, 50:325-332, 1959.

2. Barmack, J. E.: Boredom and Other Factors in the Physiology of Mental Effort: An Exploratory Study. ARCHIVES OF PSYCHOLOGY, New York, No. 218, 1937.

3. Broadbent, D. E.: Decision and Stress. London: Academic Press, Inc., 1971.

4. Buckner, D. N., A. Harabedian, and J. J. McGrath: A Study of Individual Differences in Vigilance Performance. Human Factors Research Inc., Los Angeles, California, Technical Report 2, January 1960.

5. Cattell, R. B., and H. W. Eber: Sixteen Personality Factor Questionnaire. Champaign, Illinois: Institute for Personality and Ability Testing, 1957.

6. Cattell, R. B., J. L. Horn, A. B. Sweney, and J. A. Radcliffe: Motivation Analysis Test. Champaign, Illinois: Institute for Personality and Ability Testing, 1964.

7. Claridge, G. S.: Personality and Arousal. London: Pergamon Press, 1967.

8. Coles, M. G. H., and A. Gale: Physiological Reactivity as a Predictor of Performance in a Vigilance Task. PSYCHOPHYSIOLOGY, 8:594-599, 1971.

9. Colquhoun, W. P.: Evaluation of Auditory, Visual, and Dual-Mode Displays for Prolonged Sonar Monitoring in Repeated Sessions. HUMAN FACTORS, 17:425-437, 1975.

10. Corcoran, D. W. J.: Personality and the Inverted-U Relation. BRITISH JOURNAL OF PSYCHOLOGY, 56:267-273, 1965.

11. Crider, A., and C. B. Augenbraun: Auditory Vigilance Correlates of Electrodermal Response Habituation Speed. PSYCHOPHYSIOLOGY, 12:36-40, 1975.

12. Davis, D. R., and G. R. J. Hockey: The Effects of Noise and Doubling the Signal Frequency on Individual Differences in Visual Vigilance Performance. BRITISH JOURNAL OF PSYCHOLOGY, 57:381-389, 1966.

13. Davies, D. R., and G. S. Tune: Human Vigilance Performance. New York: American Elsevier Publishing Company, Inc., 1969.

14. Eysenck, S. B. G., and H. J. Eysenck: On the Dual Nature of Extraversion. BRITISH JOURNAL OF SOCIAL AND CLINICAL PSYCHOLOGY, 2:46-55, 1963.

15. Gough, H. G.: California Psychological Inventory. Palo Alto, California: Consulting Psychologists Press, 1956.

16. Guilford, J. P., and W. S. Zimmerman: The Guilford-Zimmerman Temperament Survey. Beverly Hills, California: Sheridan Supply Company, 1949.

17. Gunn, W. J., and M. Loeb: Correlation of Performance in Detecting Visual and Auditory Signals. AMERICAN JOURNAL OF PSYCHOLOGY, 80:236-242, 1967.

18. Halcomb, C. G., and R. E. Kirk: Organismic Variables as Predictors of Vigilance Behavior. PERCEPTUAL AND MOTOR SKILLS, 21:547-552, 1965.

19. Hatfield, J. L., and M. Loeb: Sense Mode and Coupling in a Vigilance Task. PERCEPTION AND PSYCHOPHYSICS, 4:29-36, 1968.

20. Jackson, D.: The Personality Research Form. Goshen, New York: Research Psychologists Press, Inc., 1965.

21. Jenkins, H. M.: The Effect of Signal-Rate on Performance in Visual Monitoring. AMERICAN JOURNAL OF PSYCHOLOGY, 71:647-661, 1958.

22. Mackworth, J. F.: Vigilance and Habituation. Baltimore, Maryland: Penguin Books, Inc., 1970.

23. Mackworth, N. H.: Research on the Measurement of Human Performance. Medical Research Council Special Report Series No. 268, 1950.

24. McGrath, J. J., A. Harbedian, and D. N. Buckner: An Exploratory Study of the Correlates of Vigilance Performance. Human Factors Research Inc., Los Angeles, California, Technical Report 4, February 1960.

25. Parasuraman, R.: Consistency of Individual Differences in Human Vigilance Performance: An Abilities Classification Analysis. JOURNAL OF APPLIED PSYCHOLOGY, 61:486-492, 1976.

26. Siddle, D. A. T.: Vigilance Decrement and Speed of Habituation of the GSR Component of the Orienting Response. BRITISH JOURNAL OF PSYCHOLOGY, 63:191-194, 1972.

27. Stroh, C. M.: Vigilance: The Problem of Sustained Attention. New York: Pergamon Press, 1971.

28. Taylor, J. A.: A Personality Scale of Manifest Anxiety. JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY, 48:285-290, 1953.

29. Thackray, R. I., K. N. Jones, and R. M. Touchstone: Self-Estimates of Distractibility as Related to Performance Decrement on a Task Requiring Sustained Attention. ERGONOMICS, 16:141-152, 1973.

30. Thackray, R. I., K. N. Jones, and R. M. Touchstone: Personality and Physiological Correlates of Performance Decrement on a Monotonous Task Requiring Sustained Attention. BRITISH JOURNAL OF PSYCHOLOGY, 65:351-358, 1974.

31. Thackray, R. I., J. P. Bailey, and R. M. Touchstone: Physiological, Subjective, and Performance Correlates of Reported Boredom and Monotony While Performing a Simulated Radar Control Task. In R. R. Mackie (Ed.), Vigilance: Theory, Operational Performance, and Physiological Correlates, New York: Plenum Publishing Company, 1977.

32. Thackray, R. I., and R. M. Touchstone: Effects of Noise Exposure on Performance of a Simulated Radar Task. FAA Office of Aviation Medicine Report No. AM-79-24, 1979.

33. Tyler, D. M., W. L. Waag, and C. G. Halcomb: Monitoring Performance Across Sense Modes: An Individual Differences Approach. HUMAN FACTORS, 14:539-547, 1972.

34. Waag, W. L.: The Prediction of Individual Differences in Monitoring Performance. Unpublished doctoral dissertation, Texas Tech University, 1971.

35. Winer, B. J.: Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.

36. Wonderlic, E. F.: Wonderlic Personnel Test. Glencoe, Illinois: Author, 1922.

37. Zuckerman, M.: Manual and Research Report for the Sensation Seeking Scale (SSS). Newark, New Jersey: University of Delaware, 1972.