II. Prediction of Vigilance

D.A. DOBBINS, D.M. SKORDAHL, and A.A. ANDERSON, Human Factors Research Branch, The Adjutant General's Research and Development Command, U.S. Army, Washington, D.C.

• PART I of the present report described a study of trends in vigilance performance of a group of drivers during a seven hour driving shift. The present report deals with the prediction of individual differences in vigilance.

The aspect of vigilance that has been of most concern to researchers is the demonstrated reduction of signal detection proficiency by human monitors as a function of monitoring time. Many experimental studies have investigated the various parameters of vigilance tasks that influence detection performance for groups. In the course of these researches, wide individual differences in detection proficiency present even at the beginning of the monitoring period have been noted. In spite of the fact that individual differences in vigilance were recognized and attempts were made for their measurement as early as the 18th century (19), relatively little direct study has been accomplished until recently.

Although there is a great deal of evidence that individual differences in vigilance performance are reliably measured within a typical monitoring period (2, 4, 5, 9), there is little evidence that retest reliability over long intervals is sufficiently adequate to justify prediction attempts. There is even less evidence that standardized psychological measures are useful predictors of vigilance (2, 12, 15, 17). These studies have shown that correlation coefficients between psychological predictors and vigilance criteria are usually nonsignificant; but if significant, then low and sometimes disappearing on cross-validation; and that personal variables interact with task variables and thus frustrate attempts at generalized prediction.

OBJECTIVES AND METHOD

The present study was an attempt to predict vigilance performance using a greater variety of psychological predictor measures than is usually reported in vigilance research. The specific objectives were the following:

1. To examine the reliability and interrelationships between two measures of vigilance performance.
2. To determine the predictability of the vigilance criteria using a wide variety of standardized psychological tests and other measures.

Primary interest was in determining which, if any, of several well-known psychological domains hold most promise for the prediction of vigilance. The approach was empirical and no provision was made for cross-validating results. For these reasons the study is considered merely exploratory.

Samples

Subjects were 111 enlisted truck drivers furnished by the U.S. Army Transportation Corps to drive loaded vehicles around experimental highway surfaces of the Road Test project. The average age of the group was 22.7 years, the average years of formal education was 10.5 years, and 42 percent were enlistees.

Only two standards were set for the inclusion of drivers in the experimental group: (a) drivers must have been assigned to a particular driving loop for at least one month prior to vigilance testing (this standard was set to minimize the danger of divided attention between the driving task and the experimental vigilance task); and (b) during vigilance testing, each driver must have been exposed to at least 150 visual signals (5). This number of signals represents 1¼ hr of uninterrupted testing time, considered necessary to achieve minimum acceptable score stability. Not all drivers could be tested for this length of time because of apparatus unreliability. Scores derived from less than 150 signal presentations were excluded from analysis.

The group of 111 drivers was separated into two samples, day drivers (n = 51) and night drivers (n = 60), based on the previous finding that significant differences in over-all detection levels, trends, and inter-subject variability were found between day-shift and night-shift drivers. These samples were kept separate in the present study because different factors seemed operant in day and night vigilance scores. These factors might have differentially influenced prediction.

Data Collection

Predictor tests were administered to drivers before their driving at the Road Test. (Four of the 111 drivers were administered predictor tests after they had begun driving; however, their predictor testing preceded their criterion testing.) Tests were administered to drivers either at Ft. Eustis, Va., or at the AASHO Road Test site. Other predictor variables were gathered from the official service records. All predictor variables, including sources, are described by Dobbins, Tiedemann, and Skordahl (5). Predictor testing was administered to each newly assigned group of drivers continuously from August 1959 to August 1960.

Vigilance Testing was begun in August 1960 and continued through November 1960. Drivers were assembled in crew shacks before testing and were read standard instructions to the effect that they were participants in a human factors research experiment; drivers were urged to do their best on the vigilance task within the limits of driving safety. After a brief familiarization period with the Transportation Corps Vigilance Tester, drivers began their normal driving shift which consisted of circling one of five experimental driving loops in heavily loaded commercial trucks. Drivers saw light signals appear one at a time from the dashboard-mounted circular signal display units. Six of the panels covering the light sources were painted red and the remaining nine, white. The driver's task was to depress a foot-pedal when a light appeared in a red panel. Signals appearing in red panels were designated "critical" and appeared at the rate of 30 per hr; signals appearing in white panels were designated "noncritical," and appeared at the rate of 91 per hr.

Variables

Criteria.—Two vigilance criterion scores were developed for each driver. The first score was simply the percentage of critical signals detected:

$$\text{Percent Detections} = \frac{\text{Number of responses to critical signals}}{\text{Number of critical signals presented}}(100)$$

The second vigilance score was an index of errors of commission. This score included the number of responses to noncritical signals added to the number of responses to imaginary signals divided by the total number of noncritical signals presented:

$$\text{Percent False Detections} = \frac{\text{Number of responses to noncritical signals} + \text{Number of responses to imaginary signals}}{\text{Number of noncritical signals presented}}(100)$$

Most drivers were tested for two driving shifts in order to obtain an estimate of criterion reliability. For both of the preceding measures, the total vigilance score used in validity analyses was the sum of the two shift scores.

Two nonvigilance criteria were also developed. These were selected to reflect both general performance as a driver and general morale level. The primary purpose of including these variables was to examine their intercorrelations with the vigilance criteria. Investigators felt that their inclusion might help bring psychological meaningfulness to the interpretation of vigilance. There was no intrinsic interest in the predictability of either of the following two criteria by the predictor variables:

Over-all Adaptability Rating. This was a general performance measure completed by noncommissioned officer crew chiefs. The total score was a summation of points alloted to the following seven rating factors: military bearing, driver proficiency, driver dependability, extent of supervision required, promotability, interpersonal relations, and effectiveness as a team member.

TABLE 6

RELIABILITY COEFFICIENTS OF VIGILANCE CRITERIA

Criteria	Between Two Partial Shifts (avg. of 7.8 hr driving)				Within One Full Shift (7 hr driving)			
	Day₁ vs day₂ (n = 67)		Night₁ vs night₂ (n = 81)		Within day (n = 19)		Within night (n = 23)	
Percent detections	0.44	0.61[a]	0.41	0.58[a]	0.88	0.94[a]	0.78	0.88[a]
Percent false detections	0.26	0.42[a]	0.38	0.55[a]	0.77	0.87[a]	0.83	0.91[a]

[a]Augmented by Spearman-Brown formula.

Morale Inventory. This instrument is a standard U.S. Army morale measure completed by the drivers to assess attitudes toward several aspects of army life.

Predictors. — A total of 39 predictors and two reference variables were assembled. The various tests and other measures were arbitrarily grouped into eight predictor clusters. Tests were included within a cluster when they appeared to have been drawn from the same general psychological domain. The clusters are briefly described in the following section; descriptions of specific tests within each cluster have been furnished by Dobbins, Tiedemann, and Skordahl (5).

1. Physical:—Included four predictors—height, weight, visual acuity, and field of vision.
2. Psychomotor:—Included four predictors of eye-hand coordination and foot reaction time.
3. Perceptual Speed:—Included five predictor tests of the speed and accuracy of visual recognition and matching.
4. Cognitive:—Included six predictor tests of verbal, number, reasoning, spatial, and mechanical abilities.
5. Driver Aptitude:—Included five predictor scores from standardized army driver batteries, practical driver tests, and driving knowledge.
6. Personal History:—Included six measures—age, component—(enlistees vs draftees), years of education, marital status, history of grade reductions, and pre-assignment sick call rate.
7. Personality:—Included three self-report personality inventories calling for responses to a wide variety of questions about the self, others, society, authority, and other topics.
8. Attitudinal:—Included two measures of attitudes toward the specific subject of driving. One was a standardized driver attitude instrument; the other was a rating measure of essays written by drivers expressing their attitudes toward assignment at the AASHO Road Test site.
9. Reference:—Two variables were included for potential usefulness in explaining the interrelationships among preceding predictors and the vigilance criteria. These were total amount of time spent by drivers at the road test site, and the total number of miles driven.

Statistical Analyses

Because of skewed distributions, vigilance criteria were converted to normalized standard scores. Within the day and night samples, scores derived from separate testing sessions on different driving shifts were correlated to estimate between-shift reliability. Correlation coefficients, both Pearson product-moment and point-biserial as applicable, were computed between predictor and criterion variables. Validity coefficients were corrected for criteria attenuation.

TABLE 7

INTERCORRELATIONS, MEANS AND STANDARD DEVIATIONS OF CRITERIA

Sample	Criterion Variables	Mean	SD	Intercorrelations		
Night	(1) Percent detections	50.02[a]	10.18	(1)		
	(2) Percent false detections	52.60[a]	8.24	-0.25	(2)	
	(3) Over-all adaptability rating	23.92	3.76	-0.16	0.19	(3)
	(4) Morale inventory	101.68	21.04	0.20	-0.16	0.28[b]
Day	(1) Percent detections	49.08[a]	9.53	(1)		
	(2) Percent false detections	51.69[a]	7.36	-0.02	(2)[c]	
	(3) Over-all adaptability rating	24.70	2.71	0.03	-0.36	(3)
	(4) Morale inventory	98.76	25.64	-0.04	0.11	0.28[b]

[a]Normalized: mean of 50 and standard deviation of 10.
[b]Significant beyond the 0.05 level.
[c]Significant beyond the 0.01 level.

RESULTS

Table 6 summarizes the reliability analyses. In general, the augmented coefficients were low; three were below the 0.60 level set as the minimum acceptable reliability coefficients by some investigators.

The between shift coefficients were considerably lower than the within-shift coefficients reported in Part I for a smaller group of drivers. Other than the general fact that retest coefficients are usually smaller than internal consistency coefficients when computed for the same measure, two other factors may have contributed to the low reliability of between-shift coefficients. These factors were variable lengths of testing time from one driver to another, and variable lengths of time elapsing between the first and second testing sessions.

In general, there was little communality among the various measures. Percent Detections and Percent False Detections correlated only -0.25 in the night sample and -0.02 in the day sample; neither coefficient reached the 5 percent level of significance.

The only criterion variable that shared a significant degree of variance with other criteria was the Over-all Adaptability Rating, a measure of general driver performance. This rating correlated 0.28 with the Morale Inventory in both the day and night samples, and -0.36 with Percent False Detections in the day sample. Table 7 summarizes intercorrelations among the four criteria.

Only twelve validity coefficients were significant at or beyond the 5 percent level for either of two criteria in two independent samples. This relative frequency of significant coefficients was roughly at chance expectations. However, the large number of predictors with positive relationships with "good" criterion performance (percent detections) and negative relationships with "poor" criterion performance (percent false detections) suggested a low amount of valid variance in many predictors. Validity coefficients were consequently corrected for criterion attenuation. The correction was based on an assumed 0.85 reliability coefficient. This level of reliability was considered the maximum attainable under realistic testing conditions. The average level of validity was still low, however, even after the correction.

In terms of the relative number and pattern of significant coefficients, the Personality, Personal History, Driver Aptitude, and Perceptual Speed clusters, respectively, were the most promising for future research. The Cognitive, Physical, Psychomotor, and Attitudinal clusters were less promising. Table 8 summarizes the validity analyses.

The major results of this study were as follows:

1. Between-shift reliability estimates of vigilance criteria were low, ranging from 0.42 to 0.61.

TABLE 8

VALIDITY COEFFICIENTS FOR VIGILANCE CRITERIA

Predictor Variable	Percent Detections: Day (N = 51) Uncorrected	Percent Detections: Day (N = 51) Corrected[a]	Percent Detections: Night (N = 60) Uncorrected	Percent Detections: Night (N = 60) Corrected[a]	Percent False Detections: Day (N = 51) Uncorrected	Percent False Detections: Day (N = 51) Corrected[a]	Percent False Detections: Night (N = 60) Uncorrected	Percent False Detections: Night (N = 60) Corrected[a]
Physical								
Height	-07	-08	00	-01	01	01	09	12
Weight	05	05	21	25[b]	-02	-03	02	02
Visual acuity	05	05	-03	-04	-30[b]	-43[c]	12	15
Field of vision	-01	-01	19	23	-03	04	04	16
Psychomotor								
Foot reaction time	-10	-12	02	03	10	15	-23	-28[b]
Two-hand coord	15	18	-04	-05	-05	-07	-16	-19
Aiming	-05	-06	03	03	-08	-12	-15	-19
Tapping	01	01	-04	-05	00	01	-09	-11
Perceptual speed								
Attention detail	16	19	07	09	-30[b]	-43[c]	-11	-14
Army cler speed	11	13	06	08	02	02	-29[b]	-36[c]
Percept speed	07	09	14	17	-06	-08	-02	-03
Reaction to signals	08	10	15	19	-08	-11	-09	-11
Identical pictures	18	21	08	09	-11	-15	-12	-15
Cognitive								
Verbal	04	05	18	22	-02	-02	-22	-28[b]
Arith reasoning	10	11	02	03	-14	-20	-22	-28[b]
Pattern analysis	11	13	-11	-13	-13	-18	-07	-08
Mech. aptitude	17	20	-04	-04	12	17	-12	-15
Following directions	03	03	06	07	07	10	-11	-14
Spatial orient	07	09	-01	-01	-04	-06	-21	-26[b]
Driver aptitude								
Driver battery I	30[b]	36[c]	19	23	-19	-27[b]	05	06
Driver battery II	22	26	-13	-16	01	01	-23	-28[b]
Road test score	10	11	-16	-19	-14	-21	-05	-06
Automotive info	08	09	02	02	-08	-12	04	05
Driving know-how	06	07	-06	-07	-16	-22	-09	-12
Personal history								
Age	21	25	21	25[b]	-03	-04	22	27[b]
Component[d]	-31[b]	-36[c]	-11	-13	-08	-11	-08	-10
Years of education	27[b]	32[b]	12	15	-32[b]	-46[c]	-02	-03
Martial status[d]	-08	-09	11	13	03	05	17	21
Pre-AASHO gr red[d]	15	18	01	01	-27[b]	-39[c]	45[c]	56[c]
Pre-AASHO sick call rate	01	01	16	19	13	18	09	11
Personality								
ASDB (transport)	16	19	14	17	-38[c]	-55[c]	-01	-01
Gen adj key (ADAS-7)	33[b]	38[c]	08	10	-14	-21	-08	-10
Mech key (ADAS-7)	00	00	26[b]	32[c]	-20	-29[b]	-01	-01
Attitudinal								
Compet. speed	10	12	12	15	-07	-10	13	16
Other users rdwy	16	19	08	09	-12	-19	18	23
Cops	-07	-08	-10	-12	-17	-24	04	05
Vehicle	-14	-16	13	16	-04	-06	05	07
Over-all	-03	-04	13	15	-15	-21	02	03
AASHO asgmt	12	14	12	15	-21	-29[b]	-02	-03
Reference								
Time spent at AASHO	02	03	-14	-17	04	05	-04	-05
Total mi driven AASHO	-10	-12	15	13	00	00	-17	-21

[a]Coefficients corrected for criterion attenuation, correction made assumed a reliability coefficient of 0.85.
[b]Significant at 0.05 level.
[c]Significant at 0.01 level.
[d]Point-biserial coefficients.

2. Intercorrelations between the two vigilance measures were low and not statistically significant.

3. The morale and general driver performance criteria showed low relationships with the vigilance criteria. Rated performance as a driver was related positively to driver morale.

4. The relative frequency of significant validity coefficients between the 39 predictors and vigilance criteria in two independent samples was at chance levels. When corrected for low criterion reliability, the average level of validity remained low.

5. In terms of promise in future validation studies, the Personality, Personal History, Driver Aptitude, and Perceptual Speed predictors seemed relatively more useful. The Cognitive, Physical, Psychomotor, and Attitudinal predictors seemed less useful.

The low between-shift reliabilities obtained in the present study suggest the need for a systematic study of retest reliability of vigilance scores over longer intervals of time than has yet been accomplished. Such a study would not only add to the present knowledge of the stability of the vigilance phenomenon but would have major implications for the appropriate amount of effort to be expended in the future development of vigilance predictors.

The lack of relationship between the two measures of vigilance has some significance for the selection of monitors providing the lack of relationship is generalized to other types of vigilance tasks. These results suggest that the behaviors involved in failing to detect critical signals and those involved in responding to false signals may have quite different psychological bases. It further suggests, in future attempts to predict vigilance performance, that the opportunities for occurrence of both errors of omission and commission be represented in predictor variables. Furthermore, these two types of errors should be properly weighted in the criterion measures by their relative consequences for the specific vigilance task involved.

The relatively greater success of the Personality and Personal History predictors suggests the presence of a motivational component in the vigilance criteria. This interpretation must be a tentative one, however, inasmuch as these relationships may be due to specific motivating conditions associated with driving at the Road Test and not related to general principles of vigilance.

If one could accept the assumption that the vigilance scores derived from the experimental task administered in this study were good estimates of the characteristic amount of vigilance shown by drivers in their normal driving duties, then the results indicate that the present U. S. Army operational driver batteries predict driver vigilance as well as any of the experimental tests tried.

In summary, the present study and others into the prediction of vigilance performance seems to be that in spite of large and fairly stable individual differences in detection proficiency, the highly specific nature of the criterion along with possible subject-task interactions serve to restrict the predictive utility of standardized psychological tests and measures. The present study indicates that the best practical test battery that could be assembled from the large number of generalized predictors used would be of marginal usefulness even though highly reliable criteria and very low selection ratios were possible. Thus, although the present study has failed to add materially to the establishment of valid correlates of vigilance performance, future study may show that specifically developed predictors closely approximating the criterion task in terms of relevant signal parameters (signal rates, intensities, intersignal intervals, and sensory modes) may result in improved prediction. Refined self-report inventories and biographical predictors may also prove useful.

REFERENCES

1. Adams, O.S., and Chiles, W.D., "Prolonged Human Performance as a Function of the Work-Rest Cycle." Human Factors Dep. Report ORD 273, Lockheed Aircraft Corp. (1961).
2. Bakan, P., "Extroversion-Introversion and Improvement in an Auditory Vigilance Task." Brit. J. Psychol., 50:325-332 (1959).

3. Broadbent, D. E., "Perception and Communication." Pergamon Press (1958).
4. Buckner, D.N., Harabedian, A., and McGrath, J.J., "Human Factors Problems in Anti-Submarine Warfare: A Study of Individual Differences in Vigilance Performance." Human Factors Research, Inc., Tech. Rept. 2, NONR 2649(00)(1960).
5. Dobbins, D.A., Tiedemann, J.G., and Skordahl, D.M., "Field Study of Vigilance under Highway Driving Conditions." Human Factors Research Branch, Tech. Research Note 118 (1961).
6. Edwards, A. L., "Experimental Design in Psychological Research." Rinehart (1960).
7. Garvey, W.D., Taylor, F.V., and Newlin, E.P., "The Use of 'Artificial Signals' to Enhance Monitoring Performance." USN. Res. Lab. Dep., No. 5269 (1959).
8. Hebb, D.O., "Drives and the Conceptual Nervous System." Psychol. Rev., 62:243-253 (1955).
9. Jenkens, H.M., "The Effect of Signal-Rate on Performance in Visual Monitoring." Am. J. Psychol., 71:647-661 (1958).
10. Jerison, H.J., and Wallis, R.A., "Experiments on Vigilance: One-Clock and Three-Clock Monitoring." USAFWADC Tech. Rep., 57-206 (1957).
11. Jerison, H.J., and Wing, S., "Effects of Noise and Fatigue on a Complex Vigilance Task." WADC Tech. Rept., 57-14 (1957).
12. Kappauf, W.E., and Powe, W.E., "Performance Decrement on an Audio-Visual Checking Task." J. Exp. Psychol., 57:49-56 (1959).
13. Loeb, M., and Jentheau, G., "The Influence of Noxious Environmental Stimuli on Vigilance." J. Appl. Psychol., 42:47-49 (1958).
14. McGrath, J.J., Harabedian, A., and Buckner, D.N., "Review and Critique of the Literature on Vigilance Performance." Human Factors Research, Inc., Tech. Rept. 1, NONR 2649(00) NR 153-199 (1959).
15. McGrath, J.J., Harabedian, A., and Buckner, D.N., "Human Factors Problems in Anti-Submarine Warfare: An Exploratory Study of the Correlates of Vigilance Performance." Human Factors Research, Inc., Tech. Rept. 4, NONR 2649(00)(1960).
16. McGrath, J.J., and Hatcher, J.F., "The Effects of Irrelevant Environmental Stimulation on Vigilance Performance." Human Factors Research, Inc., Tech. Rept. 6, NONR 2649(00) NR 153-199 (1960).
17. McGrath, J.J., "Human Factors Problems in Anti-Submarine Warfare: Cross Validation of Some Correlates of Vigilance Performance." Human Factors Research, Inc., Supplement, Tech. Rept. 4, (1961).
18. McNemar, Q., "Psychological Statistics." Wiley, 2nd ed. (1955).
19. Ramul, K., "The Problem of Measurement in the Psychology of the Eighteenth Century." Amer. Psychologist, 15:256-265 (1960).
20. Ray, J.T., Martin, O.E., and Alluisi, E.A., "Human Performance as a Function of the Work-Rest Cycle." Nat. Academy of Sciences—Nat. Research Council Pub. 882 (1961).
21. Uhlaner, J.E., "The Driver Vigilance Tester." Unpublished memo to Subcommittee on Human Reactions of the Special Studies Panel of the AASHO Road Test. USA TAGO (Feb. 1960).