2013

The Psychomotor Vigilance Test: Sources of State and Trait Variance

Hans-Juergen Hoermann

THE PSYCHOMOTOR VIGILANCE TEST: SOURCES OF STATE AND TRAIT VARIANCE

Hans-Juergen Hoermann, PhD, FRAeS
German Aerospace Center (DLR)
Institute of Aerospace Medicine
Department of Aviation and Space Psychology, Hamburg/Germany

Within the context of pilot and air traffic controller selection tests the Psychomotor Vigilance Test (PVT, Dinges & Powell, 1985) was evaluated for its underlying sources of variance. The PVT is a simple visual reaction time task, which is well established as a measure of alertness during sustained operations. It provides scores for mean reaction times and number of lapses. While the PVT has proven sensitivity for temporary states of fatigue and other stressors in within-subjects designs, validation studies are lacking to examine it for potential sources of trait variance, which could lead to confounding effects. This paper presents results from a validation study of the PVT with N = 247 air-traffic controller applicants. The PVT was administered in the morning before and in the evening after the selection tests. Lapses and mean reaction times show a different pattern of stability coefficients and inter-correlations with other tests. PVT-lapses appear to be more sensitive to state changes than mean reaction times. However, the change scores of the lapses seem to be confounded with individual personality traits. The PVT reaction time scores show significant correlations to selection tests of psychomotor skills, response orientation and vigilance. Implications for using the PVT as a potential selection test for pilots or air-traffic controllers are discussed.

The Psychomotor Vigilance Test (PVT, Dinges & Powell, 1985) has been widely used in sleep deprivation experiments to assess the impact of fatigue on performance (Elmenhorst et al., 2009). The PVT is a visual reaction time task asking subjects to respond as soon as possible to simple stimuli usually presented on a computer screen. The standard test length is 10 minutes with about 60 to 70 signals.

The test has proven empirical validity and sensitivity as a measure of effects of fatigue and other stressors (e.g. Loh et al., 2004; Elmenhorst et al., 2009; Basner & Dinges, 2011). Furthermore, it was stated that the PVT is "ideal" for repeated use in within-subjects designs because of its short testing time and independence of learning effects and aptitude differences, which often affect other cognitive tests (Dorrian, Rogers & Dinges, 2005; Dinges et al., 2012). However, studies are lacking, which examine the PVT for potential sources of trait variance that could result in confounding effects.

The goal of this study is to examine the temporal stability of different PVT scores and to examine the PVT for potential underlying sources of trait variance in the context of standard aptitude tests for air-traffic controller selection. Additionally, interindividual differences with respect to the subjects' vulnerability to effects of test fatigue and time of day are assessed.

Methods

Subjects

N = 247 applicants (88 or 35% female) for the air-traffic controller (ATCo) ab-initio training of the DFS in Germany participated in the study. For 207 subjects complete data sets were available with two points of measurement for the PVT and the Subjective Fatigue Checkcard (Samn & Perelli, 1982). The additional 40 subjects had only one point of measurement. The subject's age range was 18 to 25 with an average of 19.8 years.

Measurements

Psychomotor Vigilance Test (PVT). A ten-minute version of the PVT was administered. Different performance scores were calculated:

- Overall mean reaction times ($PVT_{(1/2)}$-MRT)
- Mean reaction times for the 10% fastest responses ($PVT_{1/2}$-RTF)
- Mean reaction times for the 10% slowest responses ($PVT_{1/2}$-RTS)
- Number of lapses with RTs > 400ms ($PVT_{1/2}$-LAP)

The indexes "$_{1/2}$" refer to the time of measurement (PVT_1 stands for an evening measurement and PVT_2 for the morning measurement). In all calculations of the PVT scores responses within the first minute were skipped in order to compensate for potential effects of adaptation and repetition.

Subjective Fatigue Checkcard (FAT). With the $FAT_{1/2}$ (Samn & Perelli, 1982) the subjective level of fatigue was assessed subsequent to the PVT on day 1 and prior to the tests on day 2.

Performance Tests. All subjects were participants in the first stage of ab-initio selection for ATCos at the DLR (Eissfeldt, 1998). From this data set four typical aptitude tests were chosen to validate the PVT scores:

- Mental Concentration Test (MCT): The MCT involves a combination of different cognitive functions such as visual search, working memory, decoding speed, and simple arithmetic under time pressure.
- Visual Perceptual Speed (VPS): The VPS measures the ability to quickly grasp certain details of visually presented information under limited time.
- Vigilance Test (VIG): The VIG measures sustained attention for monotonous visual and auditory stimuli over a longer time period.
- Choice Reaction Time (CRT): The CRT measures mean reaction times to complex visual stimuli.

Personality. In addition, three performance related personality scales were included in the analysis. The three scales were chosen from DLR's personality questionnaire TSS (Goeters et al., 1993):

- Achievement Motivation (ACH): being ambitious, hardworking, competing
- Rigidity (RIG): being orderly, correctly, punctual, and conscientious
- Vitality (VIT): being active in sports, setting store by fitness, robust

Procedure

The selection tests were administered on two consecutive days. Day 1 has eight hours of performance testing (8am to 5pm) plus one hour lunch break. On day 2 the personality test was administered. 207 subjects completed the PVT and the FAT twice: First, between 5 and 6pm after the regular selection tests had been completed on day 1. Second, at about 9am in the morning of day 2 prior to the personality questionnaire. In this paper, the scores of PVT and FAT are indexed with the time of measurement (1 = evening of day 1; 2 = morning of day 2).

Results

Test-retest correlations were calculated for the PVT and FAT to illustrate the stability of the different scores from the evening after the selection tests to the next morning after one full night of rest (see Table 1). PVT RT measures seem to be more stable than lapses and the subjective fatigue score.

Table 1.
*Test-retest correlations for PVT and FAT from day 1 to day 2 (N = 207). (*p < .05; **p < .01).*

Test	Score	r_{tt}
PVT		
	MRT	.53**
	RTF	.61**
	RTS	.46**
	LAP	.42**
FAT		
	Total	.18**

In order to assess the effect of time of day plus being fatigued after eight hours of testing, scores of PVT and FAT were analysed by t-tests for dependent samples. All effects indicate a significant increase of alertness from evening of day 1 evening to morning of day 2 (see Table 2). Effect sizes are largest for the Fatigue Checkcard followed by PVT lapses.

Table 2.
*Effects of time of day on PVT and FAT scores. Results of t-tests with two-tailed significance levels (*p < .05; **p < .01, N = 207). Effect sizes d were calculated according to Morris & DeShon (2002).*

Test	Score	Day 1	Day 2	d	T
PVT					
	MRT	248.5	244.5	0.18	2.54*
	RTF	204.3	201.3	0.17	2.92**
	RTS	306.3	299.4	0.22	2.95**
	LAP	4.5	3.4	0.31	4.09**
FAT					
	Total	11.2	6.9	1.64	16.18**

In Table 3 the intercorrelations of PVT and FAT scores with the performance tests of the ATCo selection are shown. Correlations are higher for day 1 than for day 2 and again PVT lapses and subjective fatigue show the lowest values.

Table 3.
*Intercorrelations of PVT and FAT scores with performance tests of ATCo selection. N = 247 for day 1 and N = 207 for day 2. (*p < .05; **p < .01).*

Test		MCT	VPS	VIG	CRT	$FAT_{1/2}$
PVT	Score:					
	MRT_1	-.22**	-.19**	-.29**	-.27**	.20**
	RTF_1	-.15*	-.13*	-.20**	-.19**	.14*
	RTS_1	-.24**	-.20**	-.34**	-.28**	.21**
	LAP_1	-.11	-.20**	-.23**	-.20**	.24**
	MRT_2	-.14*	-.16*	-.18**	-.28**	.06
	RTF_2	-.13	-.09	-.06	-.19**	.09
	RTS_2	-.04	-.16*	-.15*	-.21**	.11
	LAP_2	.02	-.11	-.11	-.16*	.00
FAT						
	$Total_1$	-.04	-.03	-.06	-.08	
	$Total_2$	-.12	-.02	.00	.09	

In Table 4 the intercorrelations of PVT and FAT scores with the TSS personality scales of the ATCo selection are shown. It is remarkable that the TSS-personality scale Vitality correlates significantly and consistently with all PVT measures except the lapses. Other correlations of PVT and FAT with TSS-scales (including Extraversion) are negligible.

Table 4.
*Intercorrelations of PVT and FAT scores with personality scales of ATCo selection. N = 207. (*p < .05; **p < .01).*

Test		ACH	RIG	VIT
PVT	Score:			
	MRT_1	.06	.05	-.28**
	RTF_1	.01	.04	-.25**
	RTS_1	.07	.09	-.28**
	LAP_1	.03	.04	-.11
	MRT_2	.04	-.01	-.18**
	RTF_2	.01	.02	-.21**
	RTS_2	.03	.04	-.22**
	LAP_2	-.02	-.15*	-.11
FAT				
	$Total_1$	.06	-.04	-.16*
	$Total_2$	-.04	-.10	.03

To examine whether PVT performance variations across time of day are confounded with trait-related interindividual differences in aptitudes, we correlated the regression residuals for the PVT and FAT from day 2 on day 1 with the performance tests and personality scales of the selection test battery. The changes of FAT due to time of day were uncorrelated to these tests. However, changes in PVT mean reaction times are slightly correlated with the Complex Reaction Time test (-.16**). Changes in PVT lapses are slightly correlated with the TSS-scale rigidity (-.20**). Subjects with higher scores in these two selection tests seem to be slightly more resilient to effects of test fatigue and time of day.

Conclusions

The stability of the PVT from one day to the next is moderate to high (r_{tt} = .42 to .61) and similar to what is reported in the literature (Roach, Dawson & Lamond, 2006). Nevertheless, PVT performance levels showed significant decrements in the evening after subjects had been tested for eight hours compared to the performance level after a full night rest period. The calculated statistics indicate that the RT-scores are more stable and less changeable than the lapses.

The PVT has convergent validities with the Subjective Fatigue Checkcard (r = .20) and with aptitude tests of Vigilance (r = .20 to .34) and Choice Reaction Times (r = .19 to .28). The PVT is also related to the personality scale Vitality (r = .11 to .28). Convergent validities are higher for the evening scores than for the morning scores, which might have been caused by less interindividual variance in the morning.

Finally, it was examined whether performance differences between evening and morning scores of the PVT are in interaction with aptitudes and personality trait measures. Small but significant interactions could be identified for psychomotor speed (correlation with CRT of r = -.16) and for the personality scale Rigidity (r = -.20).

In conclusion, it was found that the PVT scores have both trait and state variance. According to the findings the state variance is larger for the lapses while the RT measures seem to contain more trait variance. Therefore, the lapses are the better indicators of stressor effects on performance in within- and between-subjects designs. However, even the sensitivity of the lapses for such effects can vary for subjects with different levels of Rigidity. Individual vulnerability to effects of stress on performance (in this study test fatigue combined with time of day) seems to be higher for subjects with lower rigidity and with slower psychomotor speed. Therefore, according to the findings presented here, PVT scores cannot be considered as being completely free of aptitude differences.

On the other hand the consistent correlation patterns of the PVT RT measures with typical selection tests have shown that the PVT scores could contribute predictive variance to other tests of attention and concentration within the context of air-traffic controller or pilot selection. Whether a shorter five-minute version of the PVT (e.g. Roach et al., 2006) can still deliver equivalent results for these aptitudes seems worth to be examined.

References

Basner, M. & Dinges, D.F. (2011). Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. *Sleep*, *34(5),* 581-91.

Dinges D.F., Powell J.W. (1985). Microcomputer analysis of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments & Computers*, *17*, 652–655.

Dinges, D.F., Mollicone, D. & Basner, M. (2012). *Psychomotor Vigilance Self Test on the International Space Station (Reaction_Self_Test).* Experiment/Payload Overview (25th July 2012). Retrieved from http://www.nasa.gov/mission_pages/station/research/experiments/Reaction_Self_Test.html.

Dorrian J, Rogers NL, Dinges DF. (2005) Psychomotor vigilance performance: a neurocognitive assay sensitive to sleep loss. In: Kushida C (Ed.), *Sleep deprivation: clinical issues, pharmacology and sleep loss effects (pp. 39–70).* Marcel Dekker, Inc., New York, N.Y..

Eissfeldt, H. (1998) Exemplary Selection Systems - The selection of air traffic controllers. In K.M. Goeters (Ed.) a*viation psychology: A science and a profession (73-80)*. Ashgate, Aldershot, UK.

Elmenhorst D. , Elmenhorst E.-M., Luks N., Maass H., Mueller E.W., Vejvoda M., Wenzel J., Samel A. (2009). Performance impairment during four days partial sleep deprivation compared with the acute effects of alcohol and hypoxia. *Sleep Medicine*, *10*, 189–197.

Goeters, K.-M. und Timmermann, B. und Maschke, P. (1993) The construction of personality questionnaires for selection of aviation personnel. *The International Journal of Aviation Psychology*, *3*, 123-141.

Loh, S., Lamond, N., Dorrian, J., Roach, G. & Dawson, D. (2004). The validity of the psychomotor vigilance tasks of less than 10-minute duration. *Behavior Research Methods, Instruments & Computers*, *36*, 339-346.

Morris, S.B. & DeShon, R.P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*, 105-125.

Roach, G.D., Dawson, D. & Lamond, N. (2006). Can a shorter psychomotor vigilance task be used as a reasonable substitute for the ten-minute psychomotor vigilance task*? Chronobiology International, 23(6),* 1379–1387.

Samn, S.W., Perelli, L.P., 1982. *Estimating aircrew fatigue: a technique with application to airlift operations*. Brooks AFB, USAF School of Aerospace Medicine. Technical report SAM-TR-82-21.

Contact information

Dr. Hans-Juergen Hoermann
German Aerospace Center (DLR)
Institute of Aerospace Medicine
Sportallee 54 a
D-22335 Hamburg, Germany
hans.hoermann@dlr.de