Journal of Applied Research in Memory and Cognition 1 (2012) 27–33

Contents lists available at SciVerse ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

Effects of sleep deprivation on cognitive performance by United States Air Force pilots

Nadia Lopez[a], Fred H. Previc[b], Joseph Fischer[c], Richard P. Heitz[d], Randall W. Engle[e,*]

[a] Air Force Research Laboratory, Biodynamics and Protection Division, United States
[b] Texas A& M University, United States
[c] Wyle I S & E, United States
[d] Vanderbilt University, United States
[e] School of Psychology, Georgia Institute of Technology, 654 Cherry Street, Atlanta, GA 30332-0170, United States

ARTICLE INFO

Article history:
Received 25 July 2011
Received in revised form 6 October 2011
Accepted 11 October 2011
Available online 19 October 2011

Keywords:
Fatigue
Attention
Flight
Cognition
Errors

ABSTRACT

This study examined the effects of 35 h of continuous sleep deprivation on performance in a variety of cognitive tasks as well as simulated flight. Ten United States Air Force pilots completed the Multi-Attribute Task Battery (MATB), Psychomotor Vigilance Task (PVT), and Operation Span Task (OSPAN), as well as simulated flight at 3 h intervals over a 35 h sleep deprivation period. Performance declined on all tests after about 18–20 h of continuous sleep deprivation, although the degree to which performance degraded varied. During the second half of the sleep deprivation period, performance on the simulated flight was predicted by PVT and OSPAN reasonably well but much less so by the MATB. Variance from optimal flight performance was predicted by both PVT and OSPAN but each measure added incremental validity to the prediction. The two measures together accounted for 58% of the variance in flight performance in the second half of the sleep deprivation period.

In August of 1997 a Korean Airline 747 in good working order flew into a hillside in Guam killing all 227 people aboard. The investigation concluded that the accident occurred because of reduced situation awareness on the part of the pilot leading to 'controlled flight into terrain'. The pilot was apparently unaware of the mountain in his path. But why would an experienced senior pilot make such an error? One possible reason is fatigue. As noted in one account of the accident, "Prior to the flight to Guam he had flown from Seoul to Australia, back to Seoul, then to Hong Kong, and then back to Seoul again before his fateful trip to Guam, all with only a few hours of rest." (http://www.airlinesafety.com/editorials/PilotFatigue.htm). Though this paper will focus on the aviation industry, assessment and prediction of effects of fatigue and sleep deprivation on performance of complex multi-tasks is a quite general issue.

The effects of fatigue on performance have long been a safety concern to terrestrial and air transportation, medicine, and industrial settings such as nuclear plants. Between 1974 and 1992, 25% of major mishaps involving US Air Force tactical fighter jets and 12% of US Navy major mishaps involved fatigue as a contributing factor (Ramsey & McGlohn, 1997). Fatigue played a role in over 12% of all US Air Force mishaps (Luna, 2003) and is a recurring concern throughout the aviation industry (Caldwell, 2005). Caldwell and Gilreath (2002) surveyed Army aircrew personnel, who reportedly worked an average of 65 h per week but obtained only 6.4 h of sleep per night. In the absence of sufficient sleep, response accuracy and speed degrades, and aircrews are more likely to lower their standards of performance, suffer impairments in the ability to integrate information, and experience a reduced attention span that may lead to forgetting or ignoring important aspects of flying (Perry, 1974). Therefore, it is important to better predict when personnel are likely to experience fatigue-based decrements using quick and easy-to-administer cognitive tests.

Various studies have shown significant performance decrements on basic cognitive tasks in fatigued subjects (Caldwell et al., 2003; Dinges et al., 1997; Matthews, Davies, Westerman, & Stammers, 2000). And, there is great potential for using such tasks to predict when a fatigued individual might be prone to performance decrements on real-world tasks such as flying airplanes or driving long-haul trucks. However, the extent to which simple cognitive tasks can be used to predict performance decrements on real-world tasks remains unclear. Caldwell et al. found physiological measures obtained from eye tracking and EEG are often difficult and expensive to incorporate into real-world military missions. Further, those measures do not appear to reliably predict fatigue decrements or performance on criterion tasks. Even if they were

---

* Corresponding author. Tel.: +1 404 644 8152.
E-mail address: randall.engle@gatech.edu (R.W. Engle).

2211-3681/$ – see front matter © 2011 Society for Applied Research in Memory and Cognition. Published by Elsevier Inc. All rights reserved.
doi:10.1016/j.jarmac.2011.10.002

more reliable or more valid, there is little connection between the theoretical constructs underlying those assessments and error in the real-world tasks.

However, several cognitive tasks do have promise. One is the Multi-Attribute Task Battery (MATB) (Comstock & Arnegard, 1992), a computerized aviation simulation test. This test has been considered desirable because of its complexity, face validity, and sensitivity to fatigue (Caldwell et al., 2003; Caldwell & Ramspott, 1998). MATB tracking accuracy may also have the ability to predict flight performance changes since Caldwell et al. found it significantly related to changes in overall flight performance accuracy after 37-h of continuous sleep deprivation.

A test widely used to assess fatigue in the aviation industry is the Psychomotor Vigilance Task (PVT) developed by Dinges and Powell (1985) (see also Lim & Dinges, 2008). This test typically involves a visually presented object that periodically changes and the subject's task is to rapidly indicate when each change occurs. For example, one typical object is a display of zeros either on a computer screen or wrist-worn display that suddenly begin counting up each millisecond and the subject is to press a button as quickly as possible to keep that number as small as possible. The logic behind the task is that if fatigue leads to a lapse of attention it should be reflected by slowed responding to the change in the stimulus event. In fact, performance on the PVT has been shown to reliably decline with reduced sleep (Dinges et al., 1997; Kamdar, Kaplan, Kezirian, & Dement, 2004; Russo et al., 2005; Whitmore et al., 2004). The PVT also reflects reliable individual differences in putatively rested individuals (Unsworth & Spillers, 2010). The PVT has rapidly become somewhat of a 'gold standard' for measuring fatigue in aviation studies but there still is much that we do not know about the task and what it measures. While there is considerable literature showing that detection time on the PVT degrades with sleep deprivation, there is a scarcity of work showing the relationship of those measures with performance on real-world tasks. Further, there is a dearth of published studies comparing the PVT with other potential measures of cognitive fatigue and how they predict performance on real-world tasks.

The Operation Span Task (OSPAN; Unsworth, Heitz, Schrock, & Engle, 2005) is a test which measures cognitive control aspects of working memory. While the OSPAN has been used to assess abiding trait-like aspects of cognitive control, it also reflects momentary state-like aspects of cognitive control and self-regulation (see Ilkowska & Engle, 2010, for a review). There is reason to believe that working memory performance as measured by the OSPAN may correlate well with flight performance during fatigue. General cognitive ability is the best predictor of pilot and navigator overall performance (McHenry, Hough, Toquam, Hanson, & Ashworth, 1990; Olea & Ree, 1994; Ree & Earles, 1991; Ree, Earles, & Teachout, 1994). Further, there is an extensive literature showing a strong relationship between working memory capacity and fluid intelligence (Engle & Kane, 2004; Engle, Tuholski, Laughlin, & Conway, 1999; Kane et al., 2004; Kyllonen & Christal, 1990). Whether these findings in the absence of systematic sleep deprivation are related to performance under conditions of fatigue is yet to be determined. The purposes of this study were to (1) investigate overall changes in performance in fatigued USAF pilots, and (2) compare different cognitive tests as predictors of flight performance over a period of continuous sleep deprivation.

Method

Ten male active-duty, reservist, or retired pilots for the USAF served as subjects after signing an informed consent agreement that detailed the procedures of the study. The mean age of the pilots was 34 years, ranging from 23 to 46 years old. All subjects were screened for near-vision acuity, vestibular function, and eye tracking quality prior to their participation. The subjects reported an average of 2805.7 total flight hours (ranging from 207 to 5000 h). Subjects self-reported an average of 7.46 h of sleep per night the week prior to the study. None were taking any type of medication known to impact mental alertness (i.e., sedating antihistamines, sleep medications, prescription stimulants, etc.) nor were they heavy nicotine (>1 cigarette per day) or heavy caffeine (>100 mg of caffeine per day) users. Subjects completed the study in groups of two. Subjects were compensated for their participation.

Apparatus

This study was conducted in the Aviation Sustained Operations Laboratory at Brooks City-Base, Texas. In addition to the cognitive tests and flight performance measurements described in this paper, EEG and eye tracking data were collected in the Gyroflight Sustained Operations Flight Simulator (Environmental Techtonics Corp., Southampton, PA). Methodology for EEG and eye tracking movements are discussed in detail in a companion paper (Previc et al., 2009). Flight performance measurements were collected in the flight simulator. The cognitive measures were collected on desktop computers in a sound-attenuated testing room as were two additional measures of subjective fatigue: the Profile of Mood States (POMS) (McNair, Lorr, & Droppleman, 1981) and the Visual Analog Scale (VAS) (Penetar et al., 1993). Subjective measures of fatigue are discussed in detail in Previc et al. (2009).

Multi-Attribute Task Battery (MATB)

The MATB (Comstock & Arnegard, 1992) is a computerized aviation simulation test that required subjects to perform an unstable-tracking task with a joy stick while concurrently monitoring warning lights and dials, responding to computer-generated auditory requests to adjust radio frequencies, and managing simulated fuel flow rates (using various key presses) for 30 min.

Data on tracking errors, response times, time-outs, false alarms, and accuracy rates were collected by means of the MATB processing software. However, only four MATB measures – communications reaction time, systems monitoring reaction time to dials, systems monitoring reaction time to lights, and tracking root mean square error (RMSE) – were analyzed because these measures are the only ones shown to be sensitive to fatigue in past studies (J.A. Caldwell, personal communication, 2006).

Operation Span Task (OSPAN)

On each trial of the automated version of the Operation Span Task subjects first see an arithmetic equation, then indicate whether an answer following the equation is correct, and finally they see a letter to remember for later recall (Unsworth et al., 2005). Three to seven such processing-and-storage presentations constitute a trial. After the trial, a recall grid is presented and subjects use a mouse to click on the letters they saw during the trial in correct serial order. To ensure subjects attended the math problems and did not rehearse the letters during that time, a 75% accuracy criterion was required on the math portion of the task. OSPAN produced an "absolute OSPAN score" (OSCORE) and a "total number correct score" (TSCORE). The OSCORE was the sum of letters in all sets of data for which the entire data string was recalled correctly. The TSCORE was the total number of letters recalled in the correct position, regardless of whether or not the entire set was perfectly recalled. Since a high criterion is used for the math problems the math score was not analyzed.

Psychomotor Vigilance Test

The PVT-192 (Ambulatory Monitoring, Inc., Ardsley, NY) is a portable hand-held device with a test requiring 10 min to administer. The display consisted of a blank screen and at random intervals a string of red LED digits appeared and at random intervals would begin counting up in millisecond increments. Subjects were to press one of the two buttons on the device with the preferred hand as soon as the change was detected and had 1.5 s to make the response before the display reset to all zeros. The inter-stimulus intervals varied from 1 to 10 s.

The PVT and associated software provides numerous measures. We analyzed four of these measures: Lapses (reaction time greater than 500 ms), MFRRT (mean of the reciprocal of the fastest 10% of reaction times), MRRT (mean reciprocal reaction time) and MEDRT (median reaction time). These measures were retrieved via the PVT REACT software. Lapses were analyzed because they are an easily understood measure of fatigue, while the three reaction time measures were selected because they are more normally distributed than mean reaction time. We chose to use the MRRT for the analysis of correlation with flight performance.

Flight Performance (FP)

One purpose of the study was to determine which cognitive measures best predict flight performance after sleep deprivation. The 19 min flight profile consisted of seven segments. On each segment except the 'turn to' and 'intersection with' final approach the pilot was commanded to maintain a set of previously specified control or performance parameters including airspeed, horizontal situation indicator, vertical velocity indicator, bank, and longitudinal bearing and glide slope. Root mean square errors (RMSE) from the specified parameters were calculated for each segment, expressed as a percent change from baseline, and "normalized" with a log transformation. The composite (averaged over segments) normalized FP RMSE was analyzed. It is important to note that deviation from the desired baseline would not be considered an error per se but simply a deviation from optimal performance. Further details concerning the flight performance measure can be obtained from the companion paper (Previc et al., 2009).

Procedure

Each subject completed three training/familiarity sessions on the first day of participation. Beginning at 1200 on the following day and continuing through the night and throughout the next day, each subject completed 10 testing sessions that covered the final 29 h of a 35 h period of continuous sleep deprivation.

Before training began, each subject came to the laboratory, signed an informed consent agreement and was briefed on all upcoming study procedures. The subject then filled out a short biographical/medical history questionnaire and other study paperwork and was screened for near-vision acuity and vestibular function. Subjects were then taken to the flight simulator for a familiarization flight. Details of flight simulator training can be found in Previc et al. (2009).

Subjects arrived at the lab at approximately 1800 the evening prior to actual data collection to complete 3–4 h of training. Because only one subject could be run at a time in the flight simulator, testing schedules for the two subjects in each pair were staggered by 1 h throughout the entire study. Aside from this time difference, schedules among all subjects remained the same.

Two training flights (during which subjects donned the eye tracking headpiece) were conducted the evening prior to the study run. Between Training Flights 1 and 2, subjects received instruction and training on the MATB (three 10 min blocks and two 20 min blocks) and OSPAN (one 15 min test). Once this training session was completed at approximately 2100, each subject returned home or to base quarters, refrained from caffeine and alcohol use, and was instructed to sleep eight hours. When subjects arrived the next morning (at 0730 or 0830), they received a small breakfast and another training flight. Subjects also completed one iteration of the subjective mood questionnaires, PVT (10 min), OSPAN (15 min) and MATB (30 min).

After all training was complete, EEG electrodes were attached and lunch was served (no caffeine was allowed during the study). Testing began at approximately 1200 or 1300 on Day 2, depending on which schedule the subject was on. First, resting EEG recordings were collected, then researchers fitted and calibrated the eye tracking headpiece, and then the subject began the flight profile. Flight performance, EEG and eye tracking data were recorded simultaneously and are discussed at length in Previc et al. (2009).

At the conclusion of each flight, and after another resting EEG was completed, 60 min of cognitive testing was performed in sound-proof chamber. Once all tests and questionnaires were completed, the subject was given a 1 h supervised break. This cycle was repeated until all 10 test sessions were completed at approximately 1700 or 1800 on Day 3. The complete training and testing schedule is detailed in Table 1.

Results and discussion

For all cognitive tests, a repeated measures analysis of variance (ANOVA) was performed on each outcome measure to test for changes across the 10 sessions. Because Mauchly's test for sphericity of the covariance matrix was significant in each case, Huynh–Feldt adjustments to the degrees of freedom were made for each ANOVA. When significance was obtained, post hoc *t*-tests were used to compare each time point with baseline levels. An alpha level of .05 was used as the criterion for all statistical tests.

To assess individual changes in the effects of fatigue relative to others and the effects of fatigue in a particular individual over time, two sets of correlations were calculated: between-subject and within-subject, respectively. The correlations were performed on the relationship between flight performance and one measure from each cognitive task: MATB-tracking RMSE, OSPAN OSCORE, and PVT MRRT. These measures were chosen because they have been shown to be sensitive measures in previous studies and exhibited consistent changes across the 35 h sleep deprivation period in the present study.

To compute the between-subject correlations, the change from the "rested" state (average of Flights 1–5) to the "fatigued" state (average of Flights 6–10) was first computed for each subject and measure. Using these changes, a Pearson's product-moment correlation coefficient was calculated for each pairwise combination of outcome measures. The purpose of computing these correlations was to determine whether there was a consistent relationship across subjects between each cognitive measure and simulator flight performance.

The within-subject correlations were calculated as follows: For a given subject and pair of variables (say PVT and FP), there are 10 pairs of data points representing the 10 sessions that the subject encountered over time. The correlation was calculated from these 10 pairs of points. These within-subject correlations determined whether, per individual, cognitive performance changes were related to simulator flight performance changes. That is, do the two variables tend to behave similarly over the 10 testing sessions?

Multi-Attribute Task Battery Data

Tracking RMSE was the only MATB measure to show significant changes over the sleep deprivation period, $F(5.43, 29.67) = 5.50$, $p < .05$, with tracking RMSE generally becoming larger over time (Fig. 1). Post hoc comparisons revealed tracking RMSE was significantly different from baseline levels beginning at 0100 on Day 3

Table 1
Schedule of training and testing.

Time	Task	Hours awake
Day 1		
1800 (Training)	Arrive	
1805 (Training)	ET, flight	
1845 (Training)	MATB, OSPAN	
1940 (Training)	ET, flight	
2015 (Training)	MATB	
2100	Release (sleep 8 h)	
Days 2 and 3		
0730	Arrive/breakfast	1.5
0800 (Training)	OSPAN	2
0820 (Training)	ET, flight	2.5
0905 (Training)	POMS, VAS, PVT, OSPAN, MATB, EEG application	3
1120	Lunch	5.5
1200 (Testing Session 1)	EEG, ET, flight	6
1300 (Testing Session 1)	POMS, VAS, PVT, OSPAN, MATB	7
1400	Break	8
⋮	⋮	⋮
1500 (Testing Session 10)	EEG, ET, flight	33
1600 (Testing Session 10)	POMS, VAS, PVT, OSPAN, MATB	34
1700	EEG removal, debrief, release	35

Note: MATB, Multi-Attribute Task Battery; OSPAN, Operation Span Task; ET, eye tracking; POMS, Profile of Mood States; PVT, Psychomotor Vigilance Task; EEG, electroencephalogram recordings.

Table 2
Between-subject flight performance error correlations with (95% confidence limits).

	Changes in MATB tracking RMSE	Changes in OSPAN OSCORE	Changes in PVT MRRT
Changes in flight performance error	0.22 (−.48, .75)	−0.65* (−.91, −.04)	−0.69 (−.92, −.11)

Note. MATB tracking RMSE, Multi-Attribute Task Battery tracking root mean square error; OSPAN OSCORE, Operation Span Task absolute OSPAN score; PVT MRRT, Psychomotor Vigilance Task mean reciprocal reaction time.

* Statistically significant at the .05 alpha level.

Table 3
Within-subject flight performance error correlations.

Subject #	Flight performance error – MATB tracking RMSE	Flight performance error – OSPAN OSCORE	Flight performance error – PVT MRRT
1	0.41	0.62	−0.40
2	0.70*	0.16	−0.64*
3	−0.03	0.51	−0.12
4	0.52	−0.38	−0.87**
5	0.05	−0.04	−0.09
6	0.82**	−0.65*	−0.82**
7	−0.56	−0.06	−0.03
8	−0.11	0.10	−0.02
9	−0.03	−0.63*	−0.48
10	0.50	−0.86**	−0.73*

Note. MATB tracking RMSE, Multi-Attribute Task Battery tracking root mean square error; OSPAN OSCORE, Operation Span Task absolute score; PVT MMRT, Psychomotor Vigilance Task mean reciprocal reaction time.

* Significant at $p < .05$.

** Significant at $p < .01$.

(Session 5), $p = .017$, continuing until 1000 on Day 3 (Session 8), $p = .007$.

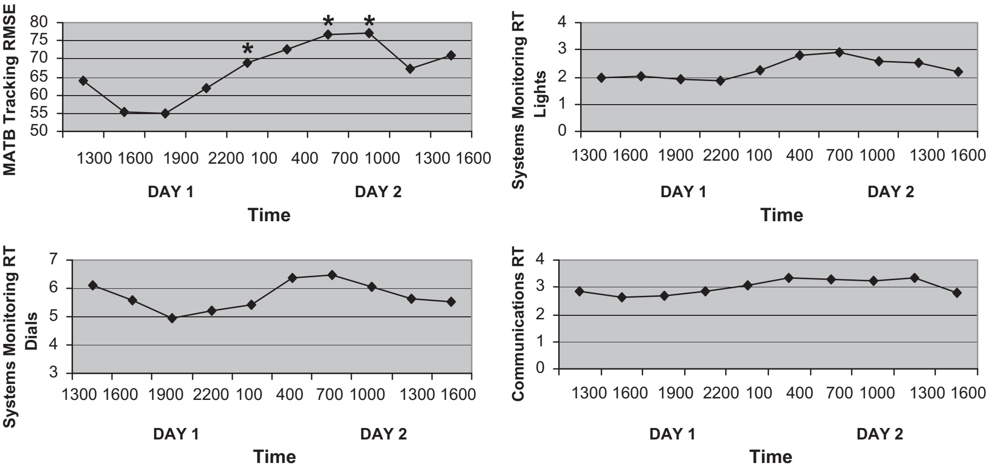
The between-subject correlation of the change in MATB tracking RMSE with the change in flight performance composite RMSE was low, $r = 0.22$ (Table 2). In addition, the within-subject correlations between MATB tracking RMSE and flight performance composite RMSE were low or in the wrong direction for 5 of the 10 subjects, $r = -0.56$ to 0.05 (Table 3). These results suggest that MATB tracking RMSE changes and flight performance changes may not be related.

Operation Span Task Data

Both OSPAN scores, OSCORE and TSCORE were significantly affected by fatigue, with subjects obtaining lower scores across time, $F(4.04, 36.38) = 2.84$, $p = .038$ and $F(3.43, 30.90) = 2.73$, $p = .054$, respectively.[1] Post hoc comparisons revealed that TSCORE was significantly lower than baseline at 0400 on Day 3 (Session 6) and both TSCORE and OSCORE were significantly lower than baseline levels at 0700 on Day 3 (Session 7), $p = .017$ and $p = .028$, respectively (Fig. 2). The between-subject correlation of the change in OSCORE with the change in FP RMSE was high, $r = -0.65$ (Table 2). Also 3 of the 10 within-subject correlations were high, $r = -0.63$ to -0.86, but four subjects showed correlations in the wrong direction (Table 3).

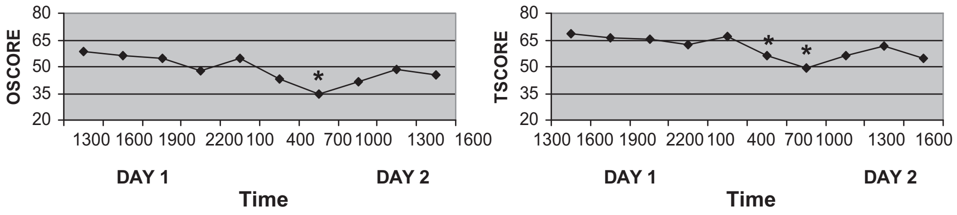
---

[1] It should be noted that only 1 of 100 data points did not meet the imposed 75% math accuracy criterion. That point was estimated using a standard statistical estimation procedure.



Note: * significantly different from the initial trial (p<.05).

Fig. 1. MATB measures across time. *Note*: *significantly different from the initial trial ($p < .05$).



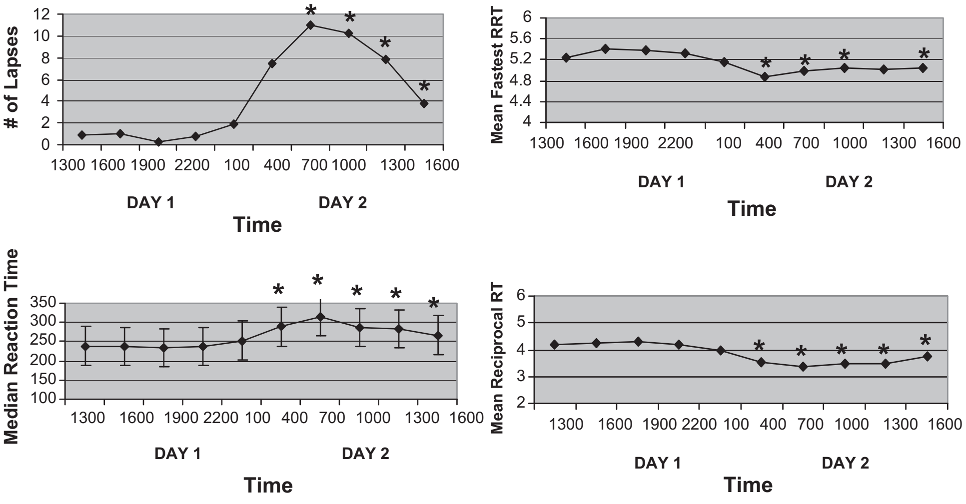
Note : * significantly different from the initial trial (p<.05).

Fig. 2. OSCORE and TSCORE scores from the Ospan task across time. *Note*: *significantly different from the initial trial ($p < .05$).

Psychomotor Vigilance Test Data

All four PVT measures were significantly affected by fatigue. The number of lapses was significantly larger as a function of time, $F(2.42, 21.80) = 5.93$, $p = .006$, and MEDRT increased significantly, $F(2.28, 20.54) = 5.71$, $p = .009$. MFRRT and MRRT were also significantly affected by fatigue, $F(7.51, 67.56) = 31.29$, $p < .001$; $F(3.35, 30.11) = 10.44$, $p < .001$ (see Fig. 3).

Post hoc comparisons showed the number of lapses became significantly larger from baseline beginning in the 0700 session (Session 7), $p = .034$, and continued throughout the duration of the study. The other three measures changed significantly from



Note: * significantly different from the initial trial (p<.05).

Fig. 3. PVT lapses, mean fastest reciprocal reaction time (MFRT), median reaction time (MRT) and mean reciprocal reaction time (MRRT) across time. *Note*: *significantly different from the initial trial ($p < .05$).

Table 4
Variance in flight performance accounted for by OSCORE and PVT MRRT.

Independent variables	Dependent variables	*R*-Square from regression
PVT MRRT, OSPAN	Flight performance RMS error	.58
PVT MRRT	Flight performance RMS error	.47
OSPAN	Flight performance RMS error	.43

Proportion of variation explained by Joint PVT and OSPAN, PVT alone, and OSPAN alone.

baseline levels earlier, beginning in the 0400 session (Session 6), MEDRT, $p = .026$; MFRRT, $p = .042$; MRRT, $p = .005$, and continuing until the end of the study.

The change in PVT MRRT was found to be a good predictor of flight performance changes under fatigue. The between-subject correlation was high, $r = -0.69$ (Table 2), 4 of 10 within-subject correlations were statistically significant, and all 10 were in the expected direction (Table 3).

Both PVT and OSPAN accounted for substantial variance in flight performance but do they reflect the same variance? Table 4 and the accompanying Venn diagram in Fig. 4 show that PVT and OSPAN combine to account for 58% of the variance in FP RMSE and, as Fig. 4 shows, 32% of the variance is common but that each of the tasks also contribute substantial incremental validity to the prediction with OSPAN contributing 11% unique variance in FP above and beyond PVT and PVT accounting for 15% above and beyond OSPAN.

To summarize, our goal was to evaluate the impact of fatigue on various measures of cognition: multi-tasking (MATB), vigilance (PVT), and working memory (OPSAN) as well as simulated flight performance (FP). We measured the performance of experienced USAF pilots on three cognitive tests over 35 h of sustained sleep deprivation and evaluated the ability of each test to predict changes in simulator flight performance.

Change in MATB tracking RMSE was significantly affected by fatigue and other measures of the test followed that trend but were not significant. Performance on the MATB began declining earlier than other cognitive measures used. These decrements lasted for the entire study, recovering slightly during the last session (probably due to the alerting effect of subjects knowing they were about to be released) but never reaching baseline levels. In this study change in MATB tracking RMSE was not a good predictor of flight performance across pilots. Because of the small sample size, it cannot be firmly concluded that there is no relationship between MATB tracking RMSE changes and flight performance but these results suggest the relationship may be weak.

OSPAN performance was significantly affected by fatigue. Additionally, changes in OSPAN correlated well with changes in flight performance, with OSCORE decreasing as flight performance error increased. OSCORE accounted for 42% of the variance in flight performance decrements. The high between-subject correlation found in this study demonstrates OSPAN's ability to predict which individuals will likely suffer flight performance decrements under fatigued conditions. In addition, OSPAN has the benefit of a much shorter training time than MATB (~30 min) and correlates much better with fatigue-induced decrements in performance across subjects than the MATB.

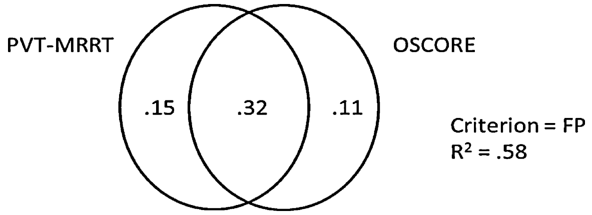


Fig. 4. Venn diagrams showing amount of variance in flight performance (RMSE) accounted for independently by OSCORE, PVT MRRT, and the two variables jointly.

Performance on the PVT declined significantly throughout this study. One measure (MRRT) accounted for 48% of the variance in flight performance over the sleep deprivation period. These results confirm previous findings (Dinges et al., 1997; Kamdar et al., 2004; Russo et al., 2005; Whitmore et al., 2004) that the PVT is a good test for measuring cognitive decrements during sustained sleep deprivation. The number of lapses between Session 4 and Session 5 jumped 37% and continued increasing after that. Not surprisingly, given the good correlation (both between- and within-subject) between changes in PVT MRRT and flight performance RMSE, decrements in flight performance began occurring at the same time. In addition to showing a significant overall fatigue effect, adequate between-subject correlations, and good within-subject correlations, the PVT is easy to use and requires very little training.

OSPAN and PVT jointly accounted for 58% of the variance in flight performance making them prime candidates for online assessment of critical fatigue points during the performance of complex tasks where less than optimal performance could be dangerous. To put this in perspective, almost 60% of the variance in deviation from optimal performance of highly experienced and skilled pilots on the one of the most complicated and sophisticated flight simulators available was predicted by two simple activities: (a) recalling a string of letters while doing simple arithmetic and (b) noticing a change in a visual display. Further, the two measures each contributed unique variance in deviation from optimal performance. Thus, if the goal is to predict performance on a complex task after sleep deprivation, a combination of the two measures would be a good place to start. For any of these measures to be used in real-world settings, we must be able to show that prediction of performance can reliably occur at the individual level. The within-subject correlations were quite noisy but those between the PVT and FP were at least all in the predicted direction. Clearly more work will be necessary before either of these measures should be used to evaluate fatigue status in the work place. Another clear limitation of the study is that sample size was limited due to access to the special subjects in this population and access to the Air Force lab facilities. That needs to be considered when interpreting our results. However, performance in the simulator was a highly practiced behavior by very skilled and experienced individuals and even with the small sample we were able to significantly detect deviation from optimal performance. Further, subjects had lots of practice on the cognitive assessments and even with that practice those assessments significantly predicted flight performance.

Practical relevance

The number of serious accidents resulting from operator error is large and seems to be growing. One way to mitigate those accidents is to have an early detection of fatigued operators and use that signal to notify the operator and others that a precarious situation exists. The PVT is rapidly becoming the 'gold standard' in the aviation industry and the present work suggests that that task along with tasks such as the Ospan might provide reliable and valid measurement of fatigue.

Acknowledgements

We wish to thank Col. Dave Tubbs of the Advanced Instrument School at Randolph Air Force Base for allowing us to recruit from his group and the pilots who participated in this research study. This research study was funded by the Air Force Research Laboratory.

References

Caldwell, J. A. (2005). Fatigue in aviation. *Travel Medicine and Infectious Disease*, *3*, 85–96. doi:10.1016/j.tmaid.2004.07.008

Caldwell, J., Caldwell, J. L., Brown, D., Symthe, N., Smith, J., Mylar, J., et al. (2003). *The effects of 37 hours of continuous sleep deprivation on the physiological arousal, cognitive performance, self-reported mood and simulator flight performance of F-117A pilots (No. AFRL-HE-BR-TR-2003-0086)*. Brooks City-Base, TX: U.S. Air Force Research Laboratory.

Caldwell, J. A. & Gilreath, S. R. (2002). A survey of aircrew fatigue in a sample of U.S. Army aviation personnel. *Aviation, Space, and Environmental Medicine*, *73*, 472–480.

Caldwell, J. A. & Ramspott, S. (1998). Effects of task duration on sensitivity to sleep-deprivation using the multi-attribute task battery. *Behavior Research Methods, Instruments, and Computers*, *30*, 651–660.

Comstock, J. R. & Arnegard, R. J. (1992). *The Multi-Attribute Task Battery for human operator workload and strategic behavior research (No. 93-5)*. Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory.

Dinges, D. F., Pack, F., Williams, K., Gillen, K. A., Powell, J. W., Ott, G. E., et al. (1997). Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4–5 hours per night. *Sleep*, *20*, 267–277.

Dinges, D. F. & Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments, and Computers*, *17*, 652–655. doi:10.3758/BF03200977

Engle, R. W. & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The Psychology of Learning and Motivation* (pp. 145–199). NY: Elsevier.

Engle, R. W., Tuholski, S. W., Laughlin, J. E. & Conway, A. R. A. (1999). Working memory, short-term memory and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General*, *128*, 309–331. doi:10.1037/0096-3445.128.3.309

Ilkowska, M. & Engle, R. W. (2010). Trait and state differences in working memory capacity. In A. Gruszka, G. Matthews, & B. Szymura (Eds.), *Handbook of individual differences in cognition: Attention, memory, and executive control, The Springer series on human exceptionality* (pp. 295–320). New York, NY, US: Springer Science + Business Media. doi:10.1007/978-1-4419-1210-7_18

Kamdar, B. B., Kaplan, K. A., Kezirian, E. J. & Dement, W. C. (2004). The impact of extended sleep on daytime alertness, vigilance, and mood. *Sleep Medicine*, *5*, 441–448. doi:10.1016/j.sleep.2004.05.003

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W. & Engle, R. W. (2004). The generality of working memory capacity: A latent variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*, 189–217. doi:10.1037/0096-3445.133.2.189

Kyllonen, P. C. & Christal, R. E. (1990). Reasoning ability is (little more than) working memory capacity?!. *Intelligence*, *14*, 389–433. doi:10.1037/0096-3445.132.1.47

Lim, J. & Dinges, D. F. (2008). Sleep deprivation and vigilant attention. *Annals of the New York Academy of Science*, *1129*, 305–322. doi:10.1196/annals.1417.002

Luna, T. (2003). Fatigue in context: USAF mishap experience. *Aviation, Space, and Environmental Medicine*, *74*, 388.

Matthews, G., Davies, D. R., Westerman, S. J. & Stammers, R. B. (2000). *Human performance: Cognition, stress and individual differences*. Philadelphia, PA: Taylor & Francis.

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology*, *43*, 335–354. doi:10.1111/j.1744-6570.1990.tb01562.x

McNair, D. M., Lorr, M. & Droppleman, L. F. (1981). *Manual for the profile of mood states*. San Diego, CA: Educational and Industrial Testing Service.

Olea, M. M. & Ree, M. J. (1994). Predicting pilot and navigator criteria: Not much more than *g*. *Journal of Applied Psychology*, *79*, 845–851. doi:10.1037/0021-9010.79.6.845

Penetar, D., McCann, U., Thorne, D., Kamimori, G., Galinski, C., Sing, H., et al. (1993). Caffeine reversal of sleep deprivation effects on alertness and mood. *Psychopharmacology*, *112*, 359–365. doi:10.1007/BF02244933

Perry, I. C. (1974). *Helicopter aircrew fatigue (AGARD Advisory Report No. 69)*. Neuilly sur Seine, France: Advisory Group for Aerospace Research and Development.

Previc, F. H., Lopez, N., Ercoline, W. R., DaLuz, C. M., Workman, A. J., Evans, R. H., et al. (2009). The effects of sleep deprivation on flight performance, instrument scanning and physiological arousal in United States Air Force pilots. *The International Journal of Aviation Psychology*, *19*, 326–346.

Ramsey, C. S. & McGlohn, S. E. (1997). Zolpidem as a fatigue countermeasure. *Aviation, Space, and Environmental Medicine*, *68*, 926–931.

Ree, M. J. & Earles, J. A. (1991). Predicting training success: Not much more than g. *Personnel Psychology*, *44*, 321–332. doi:10.1111/j.1744-6570.1991.tb00961.x

Ree, M. J., Earles, J. A. & Teachout, M. (1994). Predicting job performance: Not much more than *g*. *Journal of Applied Psychology*, *79*, 518–524. doi:10.1037/0021-9010.79.4.518

Russo, M. B., Kendall, A. P., Johnson, D. E., Sing, H. C., Thorne, D. R., Escolas, S. M., et al. (2005). Visual perception, psychomotor performance, and complex motor performance during an overnight air refueling simulated flight. *Aviation, Space, and Environmental Medicine*, *76*, C92–C103.

Unsworth, N., Heitz, R. P., Schrock, J. C. & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498–505. doi:10.3758/BF03192720

Unsworth, N. & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language*, *62*, 392–406. doi:10.1016/j.jml.2010.02.001

Whitmore, J., Doan, B., Fischer, J., French, J., Heintz, T., Hickey, P., et al. (2004). *The efficacy of modafinil as an operational fatigue countermeasure over several days of reduced sleep during a simulated escape and evasion scenario (No. AFRL-HE-BR-R-2004-0021)*. Brooks City-Base, TX: U.S. Air Force Research Laboratory.