

Psychomotor vigilance performance predicted by Epworth Sleepiness Scale scores in an operational setting with the United States Navy

NITA LEWIS SHATTUCK and PANAGIOTIS MATSANGAS

Operations Research Department, Naval Postgraduate School, Monterey, CA, USA

Keywords

fatigue, fitness-for-duty, operational performance, shiftwork, sleep deprivation

Correspondence

Nita Lewis Shattuck, PhD, 1411 Cunningham Drive, Operations Research Department, Naval Postgraduate School, Monterey, CA 93943, USA.

Tel.: (831) 656-2281;

fax: (831) 656-2595;

e-mail: nlshattu@nps.edu

Accepted in revised form 10 August 2014;
received 26 May 2014

DOI: 10.1111/jsr.12243

SUMMARY

It is critical in operational environments to identify individuals who are at higher risk of psychomotor performance impairments. This study assesses the utility of the Epworth Sleepiness Scale for predicting degraded psychomotor vigilance performance in an operational environment. Active duty crewmembers of a USA Navy destroyer ($N = 69$, age 21–54 years) completed the Epworth Sleepiness Scale at the beginning of the data collection period. Participants wore actigraphs and completed sleep diaries for 11 days. Psychomotor vigilance tests were administered throughout the data collection period using a 3-min version of the psychomotor vigilance test on the actigraphs. Crewmembers with elevated scores on the Epworth Sleepiness Scale (i.e. Epworth Sleepiness Scale >10) had 60% slower reaction times on average, and experienced at least 60% more lapses and false starts compared with individuals with normal Epworth Sleepiness Scale scores (i.e. Epworth Sleepiness Scale ≤ 10). Epworth Sleepiness Scale scores were correlated with daily time in bed ($P < 0.01$), sleep ($P < 0.05$), mean reaction time ($P < 0.001$), response speed $1/\text{reaction time}$ ($P < 0.05$), slowest 10% of response speed ($P < 0.001$), lapses ($P < 0.01$), and the sum of lapses and false starts ($P < 0.001$). In this chronically sleep-deprived population, elevated Epworth Sleepiness Scale scores identified that subset of the population who experienced degraded psychomotor vigilance performance. We theorize that Epworth Sleepiness Scale scores are an indication of personal sleep debt that varies depending on one's individual sleep requirement. In the absence of direct performance metrics, we also advocate that the Epworth Sleepiness Scale can be used to determine the prevalence of excessive sleepiness (and thereby assess the risk of performance decrements).

INTRODUCTION

Occupations such as those held by US military personnel and first-response team members are characterized by high stakes, elevated stress and grave risks. These individuals are often required to make split-second decisions with little margin for error – yet their work patterns are tremendously demanding, and their sleep is often sacrificed in order to complete operational missions. While the entire team may be sleep-deprived, it is critically important to be able to identify those individuals who are at higher risk of degraded

performance and reduced alertness (a fitness-for-duty test) to ensure the highest probability of mission success.

These operational environments also pose unique challenges for the study of human performance. Especially evident in military environments, the 'real world' demands on participants' time make compliance with testing and measurement procedures difficult. While laboratory models of scientific inquiry require strict adherence to test schedules, participants in operational environments have duties that prevent them from total engagement in a testing regime. Often, these individuals decline to participate in studies that

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE 2015		2. REPORT TYPE		3. DATES COVERED 00-00-2015 to 00-00-2015	
4. TITLE AND SUBTITLE Psychomotor Vigilance Performance Predicted by Epworth Sleepiness Scale Scores in an Operational Setting with the United States Navy				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School, Operations Research Department, Monterey, CA, 93943				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

make additional requirements on their time – or their compliance is such that missing observations or limited sample sizes make statistical inference problematic. Therefore, to the extent possible, testing procedures in operational environments must be unobtrusive, simple, reliable and short. Researchers studying work and rest patterns of individuals in operational environments need a short, reliable screening tool for identifying individuals at risk of degraded cognitive performance or restricted sleep patterns. If such a tool were available, commanders could also use it to help identify fatigue problems in their unit and to select the least-fatigued candidates for missions. With this problem in mind, we evaluated the Epworth Sleepiness Scale (ESS; Johns, 1991) to determine if ESS scores were predictive of actigraphically determined sleep and psychomotor vigilance performance of participants in an operational environment.

The ESS is a self-administered instrument designed to assess levels of daytime sleepiness. Clinicians commonly use the ESS in office settings as a screening tool to identify individuals with excessive daytime sleepiness and potential sleep disorders. Using a four-point Likert scale, participants indicate their chance of dozing off or falling asleep in eight different everyday situations. Responses are scored from 0 to 3, with 0 being 'would never doze', 1 is 'slight chance of dozing', 2 is 'moderate chance of dozing' and 3 denotes a 'high chance of dozing'. The participants are instructed to rate themselves according to 'your usual way of life in recent times'. Responses are pooled to arrive at a total score ranging from 0 to 24. A total score of more than 10 on the ESS is an indication of higher than normal daytime sleepiness and suggests the need for further evaluation (Johns, 1992). The questionnaire has a high level of internal consistency as measured by Cronbach's alpha, which ranges from 0.73 to 0.88 (Johns, 1992).

Epworth Sleepiness Scale scores are influenced by multiple factors. Olson *et al.* (1998) showed that ESS scores are affected by psychological factors such as depression and anxiety. These researchers suggest that the ESS should not be used to demonstrate or exclude sleepiness as measured by the Multiple Sleep Latency Test (MSLT). Another study with patients suspected or confirmed to have obstructive sleep apnea syndrome identified a statistically significant association between ESS and self-reported sleepiness but not with mean sleep latency (SL) on the MSLT (Chervin and Aldrich, 1999). The authors suggest that the ESS scores are not an effective surrogate for the MSLT.

In contrast to the previous studies where it was suggested that objective tests like MSLT are a more reliable metric of sleep propensity (SP), Johns (1994) focused on the fact that ESS assesses average sleepiness in daily life. His work supports the idea that individual measurements of SP involve three components of variation in addition to short-term changes: the average SP, which is a general characteristic of the participant; the situational SP, which is characteristic of the situation in which SP is measured; and a third component that is specific for both participant and situation. Johns states

that Total ESS scores provide an assessment of the individual's average SP that can be measured as reliably as the mean SL in the MSLT.

Although the ESS has been used extensively in clinical research settings to identify excessive daytime sleepiness, our review failed to identify any studies that focused on the association between ESS scores and cognitive performance in operational environments. Especially in the military environment, which is characterized by performance under conditions of chronic sleep deprivation and fatigue (Miller *et al.*, 2008, 2012), the ESS may serve as a rapid screening tool to identify individuals who are at higher risk of psychomotor performance impairments. The goal of the current study was to investigate whether ESS scores can be used to differentiate between levels of psychomotor vigilance performance in individuals working in a naval environment.

MATERIALS AND METHODS

Participants

The study sample included active duty crewmembers from USS JASON DUNHAM, a US Navy Arleigh Burke-class destroyer, Bath Iron Works, Bath, Maine, USA.

Equipment

Actigraphic estimates of crewmembers' sleep were obtained using the Motionlogger Watch [Ambulatory Monitoring Inc (AMI), Ardsley, NY, USA] using the zero-crossing mode. Analysis of the actigraphic recordings was performed using Action-W version 2.7.2155, [AMI, Ardsley, NY, USA] software using the Cole–Kripke algorithm with rescoring rules and the following parameters: epoch length = 1 min, zero-crossing mode. Actigraphic recordings followed the recommendations of Standards of Practice Committee of the American Academy of Sleep Medicine (2003). Participants were instructed to wear the wrist activity monitors on their non-dominant wrist at all times of the day and night during the study period. Participants completed a daily activity log to indicate their activities in 30-min increments, to include sleep and naps during the study. This log was tailored to the specific activities of the ship's schedule.

Performance data were collected using the psychomotor vigilance test (PVT; Dinges and Powell, 1985). The PVT is a simple reaction time (RT) test where participants press a response button as soon as the stimulus appears on the screen. PVT performance is affected not only by sleep loss but also has been shown to be sensitive to circadian rhythmicity (Dinges *et al.*, 1997; Doran *et al.*, 2001; Durmer and Dinges, 2005; Jewett *et al.*, 1999; Wyatt *et al.*, 1997). The PVT has only minor learning effects. Asymptotic performance can be reached in one-three trials (Balkin *et al.*, 2000; Dinges *et al.*, 1997; Jewett *et al.*, 1999; Kribbs and Dinges, 1994; Rosekind *et al.*, 1994). The test's nominal inter-stimulus interval (ISI), defined as the period between the

last response and the appearance of the next stimulus, ranges randomly from 2 to 10 s. The standard version of the PVT has a duration of 10 min (Loh *et al.*, 2004). However, shorter versions have also been shown to assess sleep deprivation effects (Basner and Dinges, 2011; Loh *et al.*, 2004). Because operational demands prevented the use of the 10-min version in this study, we used the 3-min version of the PVT (PVT-192) that is included as an optional feature on the AMI actigraphs. The ISI ranged from 2 to 10 s. The letters 'PUSH' were backlit in red and served as the visual stimulus. RTs were displayed to the study participants in milliseconds.

Procedures

The study protocol was approved by the Naval Postgraduate School Institutional Review Board; participants provided written informed consent before enrolling in the study. The data collection occurred from 3 December to 18 December 2012 onboard a naval vessel underway in a forward-deployed area of operations. Sea state was relatively calm during the data collection period. Participants had been in their underway routine for a period of approximately 5 months before the study commenced. After enrolling in the study, participants completed a series of questionnaires, including the ESS (Johns, 1991), the Pittsburgh Sleep Quality Index (PSQI; Buysse *et al.*, 1989) and the Morningness–Eveningness Questionnaire (MEQ; Horne and Östberg, 1976). Participants were issued Motionlogger actigraphic devices and were instructed to take the PVT four times per day, before and after standing for each of two daily watch periods. They were also asked to fill out daily activity logs divided into 30-min increments to indicate how they spent each day.

Analytical approach

This study was part of a broader study conducted onboard a USN destroyer to assess the impact of watch schedule on sleep quality and psychomotor vigilance performance. The subset of data selected for this analysis was an 11-day period from 4 December to 14 December 2012. Actigraphic recordings were used to determine bedtime, wake time and sleep episode duration. These data were entered into a Microsoft Excel spreadsheet. Statistical analysis was conducted with JMP statistical software (JMP Pro 10; SAS Institute, Cary, NC, USA). Imputation was neither needed nor used. All variables underwent descriptive statistical analysis to identify anomalous entries and to determine demographic characteristics.

Average time in bed (TIB) and sleep amounts were calculated from actigraphic data by day and participant. Sleep episode duration and bedtime/wake time were derived from the actigraphic recordings and were verified by the self-reported activity logs. After verifying the bedtime and wake times, Action-W 2.7.1 was used to calculate TIB and sleep duration. Individuals with missing actigraphy data were

excluded from the analysis to avoid systematic errors. However, those individuals who had only actigraphy (i.e. their activity logs were missing) were included in the analysis. Similarly, PVT metrics were calculated to get an average PVT score for each individual over the entire study period. Non-parametric correlations (Spearman's rho) were calculated between ESS scores, average TIB and sleep duration, and average PVT metric.

Sleep analysis was based on two metrics, the average daily TIB and the average daily sleep amount per participant. Based on recommendations of Basner and Dinges (2011), PVT performance was assessed using seven different metrics: mean RT; mean response speed (1/RT); fastest 10% RT (i.e. 10th percentile of RT); slowest 10% of 1/RT (i.e. 10th percentile of 1/RT); percentage of lapses; percentage of false starts; and percentage of lapses and false starts (combined). A 500-ms threshold has been used commonly in PVT research to identify lapses, i.e. responses with a RT greater than or equal to the specified threshold (Loh *et al.*, 2004). However, a 355-ms lapse threshold has also been applied with success (Basner and Rubinstein, 2011; Basner *et al.*, 2011). For this study, we used both the 355-ms and the 500-ms lapse thresholds for the analysis of the 3-min PVT data.

RESULTS

Sixty-nine crewmembers (53 males, 16 females; age: $M = 28$ years, $SD = 6.04$) volunteered to participate in the study. Twenty-one participants were officers and 48 were enlisted personnel, with an average of 6.15 years in service ($SD = 5.34$). Fifty-seven participants (82.6%) were studied while they were working a rotating watchstanding schedule.

The average MEQ score was 49.4 ($SD = 7.22$), ranging from 35 to 66. Eight (11.6%) of the participants were identified as 'moderately morning type', 53 (76.8%) as 'neither type' and eight (11.6%) as 'moderately evening type'. The average PSQI Global score was 8.49 ($SD = 3.30$, $MD = 8$), ranging from 3 to 18. PSQI scores indicated that only five participants (8%) were 'good sleepers' (PSQI score < 5). PSQI Global scores were negatively correlated with MEQ scores (Spearman's $\rho = -0.285$, $P = 0.021$). Study participants had an average ESS Total score of 10.6 ($SD = 3.93$, $MD = 10$), ranging from 2 to 22 (Fig. 1). Not surprisingly, ESS Total scores were significantly correlated with PSQI Global scores (Spearman's $\rho = 0.519$, $P < 0.001$).

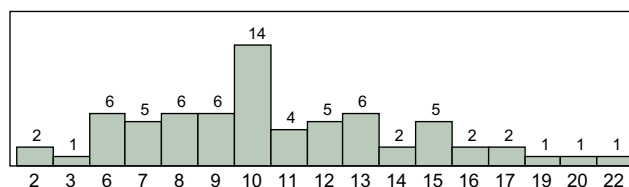


Figure 1. ESS scores. Number above each bar denotes the number of participants.

Participants received an average of 6.72 h of daily sleep (SD = 0.856, MD = 6.76), ranging from 4.9 to 8.78 h. On average, their TIB was 7.39 h daily (SD = 0.897, MD = 7.4), ranging from 5.66 to 9.65 h. Each participant had an average of 26 PVT trials (SD = 8.61, MD = 26, minimum = 12, maximum = 44). Table 1 describes PVT performance in terms of the metrics used in this study. Each of the metrics was averaged per participant.

Table 2 shows the non-parametric correlations (Spearman's rho) between ESS scores, sleep and the nine PVT metrics. Scores on the ESS were significantly correlated with daily TIB and daily sleep duration, as well as eight of the nine PVT metrics except the fastest 10% of the RTs. Both the 355-ms lapse threshold and the 500-ms lapse threshold were significantly correlated with ESS scores in a positive direction, although the 500-ms lapse threshold had a higher correlation (0.391 versus 0.453) than the 355-ms lapse threshold.

Next, participants were divided into two groups (Normal and Elevated) according to their ESS scores. The Normal ESS group was comprised of those individuals with an ESS score less than or equal to 10, while the Elevated ESS group was made up of those individuals with an ESS score greater than 10, the cutoff recommended by Johns (1991, 1992). Table 3 lists all variables that were compared (column 1), the average value and standard deviation of those variables for the Normal ESS group (column 2), the average value and standard deviation for the Elevated ESS group (column 3), the significance levels that resulted from comparing those means (column 4), and the percentage-wise difference in mean values between groups (column 5). The two-sided Wilcoxon Rank Sum test was used to calculate these differences; probabilities fluctuated from $P=0.036$ to $P<0.001$. Results showed that the two groups differ significantly in both daily TIB and sleep duration and in all of the PVT metrics assessed, except the fastest 10% RT.

The two groups differed not only in average PVT performance, but they also differed in the variability of their performance. In Table 3, column 5 shows the percent difference between the means of the two groups. The largest difference was for the variable 'Lapses over 500 ms', which

Table 2 Correlation results

Variable	ESS	Daily TIB amount	Daily sleep amount
Daily TIB amount	-0.330**		
Daily sleep amount	-0.298*		
Mean RT	0.454***	-0.278*	-0.286*
Mean 1/RT	-0.270*	0.251*	
Fastest 10% RT		-0.221	
Slowest 10% 1/RT	-0.409***	0.220	
False starts (%)	0.210		
Lapses 500 ms (%)	0.453 [†]	-0.230	-0.214
Lapses 355 ms (%)	0.391***	-0.226	
Lapses 500 ms + false starts (%)	0.485 [†]	-0.240*	-0.266*
Lapses 355 ms + false starts (%)	0.430***	-0.237*	-0.205

ESS, Epworth Sleepiness Scale; RT, reaction time; TIB, time in bed.
 Inclusion criterion: $P < 0.10$.
 * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; [†] $P < 0.0001$.

showed a 150% difference between the Normal and Elevated ESS groups. In general, crewmembers with an ESS > 10, the Elevated ESS group, had increased variability in most metrics when compared with the Normal ESS group with ESS ≤ 10.

Both daily TIB and sleep duration are significantly lower in the Elevated ESS group (also shown in Fig. 2). Compared with the Normal ESS group, the Elevated ESS group has approximately 60% slower RTs (Fig. 3) and a greater than 77% increase in false starts and lapses (Fig. 4). The Elevated ESS group also has increased variability by more than 110% in seven of the PVT metrics (mean RT, fastest 10% RT, percentage of 355-ms and 500-ms lapses, percentage of false starts, percentage of 355-ms/500-ms lapses and false starts combined) compared with the Normal ESS group. Figs 2–4 show the results of these tests for daily TIB, daily sleep duration, RT, lapses and false starts by ESS classification. The black columns indicate the Normal ESS group, while the Elevated ESS group is shown by the white column. Vertical bars represent 1 SD.

Table 1 PVT metrics for all study participants

PVT metric	M	SD	MD	Minimum	Maximum
Mean RT (ms)	396	247	336	214	1691
Mean 1/RT	3.94	0.753	3.90	0.922	5.32
Fastest 10% RT (ms)	207	80.3	192	140	780
Slowest 10% 1/RT	2.45	0.667	2.42	0.464	3.66
False starts (%)	2.23	2.19	1.69	0.15	13.0
Lapses 500 ms (%)	9.19	13.1	7.02	0.65	97.5
Lapses 355 ms (%)	18.4	15.8	14.6	1.06	99.7
Lapses 500 ms + false starts (%)	11.4	13.7	8.50	1.17	98.7
Lapses 355 ms + false starts (%)	20.7	16.2	17.4	1.49	99.5

PVT, psychomotor vigilance test; RT, reaction time.

Table 3 Comparison between Normal and Elevated ESS groups

Variable (1)	Normal group ESS ≤ 10 (2)	Elevated group ESS > 10 (3)	P-value* (4)	Percent difference in means Normal group versus Elevated group (5)	Percent difference in SD Normal group versus Elevated group (6)	P-value† (7)
Daily TIB (h)	7.59 (0.97)	7.13 (0.72)	0.032	-6.06	-25.8	0.075
Daily sleep (h)	6.91 (0.89)	6.47 (0.75)	0.036	-6.37	-15.7	0.362
Mean RT (ms)	315 (81.7)	508 (343)	<0.001	61.3	320	<0.001
Mean 1/RT	4.13 (0.6)	3.68 (0.87)	0.036	-10.9	45	0.312
Fastest 10% RT (ms)	192 (27.1)	228 (118)	0.117	18.8	335	0.023
Slowest 10% 1/RT	2.68 (0.57)	2.14 (0.68)	<0.001	-20.2	19.3	0.888
False starts (%)	1.64 (1.28)	3.04 (2.86)	0.020	85.4	123	0.004
Lapses 500 ms (%)	5.63 (4.30)	14.1 (18.6)	<0.001	150	318	0.005
Lapses 355 ms (%)	13.8 (9.53)	24.8 (20.2)	0.002	77.6	117	0.014
Lapses 500 ms + false starts (%)	7.28 (4.65)	17.1 (19.1)	<0.001	135	311	0.007
Lapses 355 ms + false starts (%)	15.4 (9.54)	27.9 (20.4)	<0.001	81.2	114	0.016

ESS, Epworth Sleepiness Scale; RT, reaction time; TIB, time in bed.

*Wilcoxon Rank Sum test results for the comparison in the mean values between groups.

†Levene's test for equality of variances between groups.

Next, we assessed the predictive ability of ESS scores to identify crewmembers with degraded PVT performance, i.e. slowed RT, or increased lapses and false starts. For this analysis, we again divided the participants into two groups. The two groups were individuals with RTs above and below the 50th percentile of all RTs. Based on this classification scheme, the odds ratio is 6.2 [95% confidence interval (CI): 2.1–18.4] for individuals with an ESS > 10 (i.e. Elevated ESS group) to have slowed RTs (i.e. RTs > 50th percentile). Similarly, the odds ratios are 7.1 (95% CI: 2.4–21.0) and 4.6 (95% CI: 1.6–13.2) for these same individuals in the Elevated ESS group to experience more 500-ms lapses and false starts or 355-ms lapses and false starts, respectively (i.e. lapse rate and false starts > 50th percentile). Stated simply, individuals with an elevated ESS score are twice as likely to experience lapses and false starts and to have slowed RTs (relative risk = 2.0–2.5).

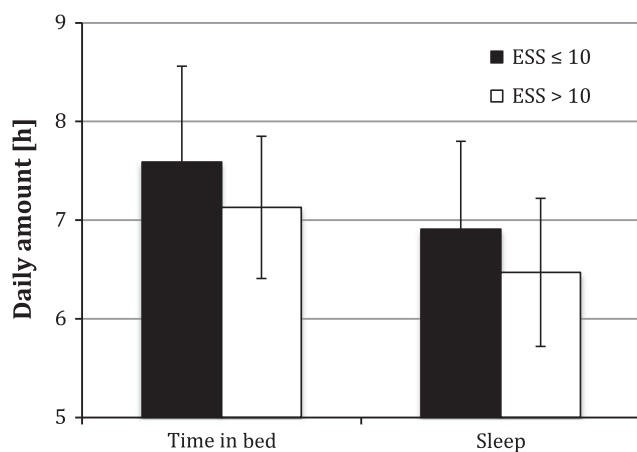


Figure 2. Average time in bed and sleep in hours by ESS classification.

DISCUSSION

Over the past decade, research into individual sleep requirements has documented that individual differences in sleep requirements exist (Grant and Van Dongen, 2013; Kuna *et al.*, 2012; Rupp *et al.*, 2012; Van Dongen, 2012). Although we do not understand the causal mechanisms, we now recognize that some individuals need less sleep than others, while other individuals appear much less vulnerable to the effects of sleep deprivation. Identifying an individual's unique sleep requirement (and determining what, if any, level of sleep debt exists at any given point in time) has proven difficult. In operational environments, this capability would be a boon to leaders who must make decisions about which team members are better rested and up to performing critical tasks.

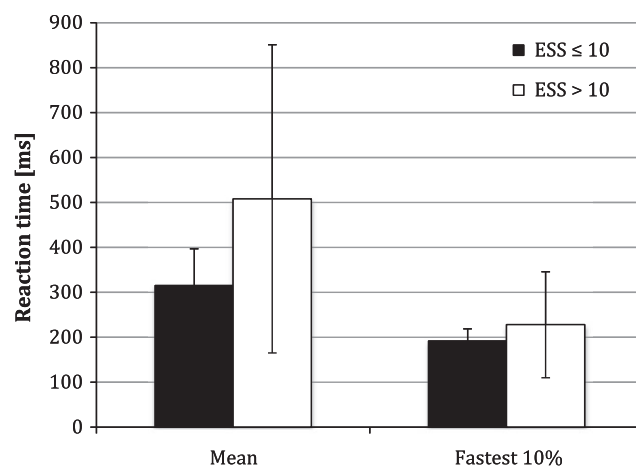


Figure 3. Average reaction time in milliseconds by participant by ESS classification.

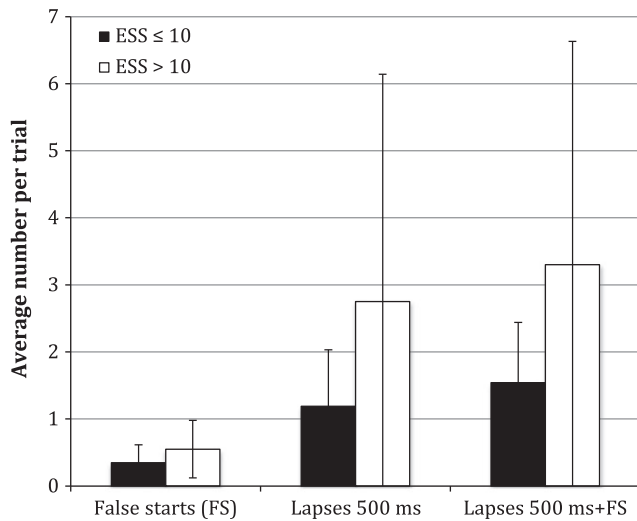


Figure 4. Frequency of false starts, lapses and lapses + false starts by ESS classification.

Results of our study showed that ESS scores were a better predictor of degraded psychomotor performance than actigraphic sleep alone. Individuals with elevated ESS scores also experienced prolonged RTs, greater numbers of lapses and false starts, and increased variability in most psychomotor performance metrics. This latter increase in variability suggests that individuals with elevated ESS scores exhibit ‘state instability’, a phenomenon in which their performance varies greatly, making it hard to rely on them to respond in a consistent manner.

If this finding can be replicated and extended to other populations, the ESS could be used as a screening tool to identify individuals who are overly fatigued and at higher risk of performance decrements – in short, a brief fitness-for-duty assessment. It is possible that the US Navy, by virtue of its selection and training process, and through the self-selected attrition of its population, has in effect screened its members and retained those individuals who are able to continue performing in the face of restricted sleep. These Navy policies may serve to weed out those who are less resistant to sleep deprivation. However, when significant levels of fatigue exist such that performance is affected, elevated ESS scores are able to identify it.

It is also possible that ESS scores would not be useful as a fitness-for-duty tool in a civilian population – or in individuals who are willing to ‘game the system’ by providing ESS answers that meet another need. There are documented situations where individuals falsify their answers either exaggerating their fatigue in order to avoid work or underestimating their fatigue in order to ensure that they are allowed to work (Parks *et al.*, 2009; Talmage *et al.*, 2008). This question of the veracity of responses is a critical issue to consider for the ESS to be used as a fitness-for-duty test. The military population may differ from the general population in its accuracy in reporting ESS scores. They may respond more honestly because inaccurate reporting could endanger

fellow service members by failing to truly report their level of sleepiness.

We theorize that ESS scores are an indication of personal sleep debt that varies depending on the current opportunity for sleep combined with an individual’s own sleep requirement. In the absence of direct performance metrics, the ESS may be a useful means of assessing fitness-for-duty by identifying individuals at higher risk of degraded psychomotor vigilance in operational environments.

Although this study focuses on the ESS, the literature provides a number of other tools measuring sleepiness (e.g. Akerstedt *et al.*, 2014; Ftouni *et al.*, 2013). Future studies should assess and compare the utility of these tools as fitness-for-duty predictors.

A limitation of this study is the extent that ESS can detect acute sleepiness. As noted by Johns (1994), the ESS assesses average sleepiness in daily life. In contrast to ESS, the PVT is sensitive to both acute (Doran *et al.*, 2001) and chronic partial sleep deprivation (Dinges *et al.*, 1997). Therefore, while our results are consistent with other data in clinical settings (Batool-Anwar *et al.*, 2014), ESS may not predict acute situational compromises in persons who are acutely sleep deprived, but not chronically so. This problem is ameliorated by the findings of multiple operational studies though, showing that military populations are chronically sleep deprived (Miller *et al.*, 2008, 2011, 2012).

In conclusion, the results from the current study suggest the potential use of the ESS in military operational environments as a simple and rapid method to identify individuals at higher risk of decrements in psychomotor vigilance performance, and for estimating the prevalence of excessive sleepiness in a given population at a point in time. Based on these findings, future studies should further investigate the generalizability of these findings to other military and operational environments.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Dr Lauren Waggoner and LT Roger Young, part of the team that participated in the data collection aboard the USS JASON DUNHAM. The authors also acknowledge the funding support of the Bureau of Navy, the Twenty-First Century Sailor Office, and the Office of Naval Personnel. Finally, most sincere thanks go to CDR David Bretz and CDR Michael Meredith, the commanding officers of the USS JASON DUNHAM, and her crewmembers who kindly consented to participate in this study despite their considerable workload and fatigue levels. Each day, you stand in harm’s way to protect our nation and make our world safer. Thank you all.

DISCLOSURE STATEMENT

This work is not an industry-supported study, and did not use off-label or investigational items. No conflicts of interest declared.

AUTHOR CONTRIBUTIONS

NLS designed the study and collected the data. PM. undertook the statistical analysis and wrote the first draft of the manuscript. Both authors contributed to the interpretation of the results and have approved the final manuscript.

REFERENCES

- Akerstedt, T., Anund, A., Axelsson, J. and Kecklund, G. Subjective sleepiness is a sensitive indicator of insufficient sleep and impaired waking function. *J. Sleep Res.*, 2014, 23: 240–252.
- Balkin, T., Thorne, D., Sing, H. *et al.* *Effects of Sleep Schedules on Commercial Motor Vehicle Driver Performance*. U.S. Department of Transportation, Springfield, VA, 2000: 446.
- Basner, M. and Dinges, D. F. Maximizing sensitivity of the Psychomotor Vigilance Test (PVT) to sleep loss. *Sleep*, 2011, 34: 581–591.
- Basner, M. and Rubinstein, J. Fitness for duty: a 3-minute version of the Psychomotor Vigilance Test predicts fatigue-related declines in luggage-screening performance. *J. Occup. Environ. Med.*, 2011, 53: 1146–1154.
- Basner, M., Mollicone, D. and Dinges, D. F. Validity and sensitivity of a brief psychomotor vigilance test (PVT-B) to total and partial sleep deprivation. *Acta Astronaut.*, 2011, 69: 949–959.
- Batool-Anwar, S., Kales, S. N., Patel, S. R., Varvarigou, V., Deyoung, P. N. and Malhotra, A. Obstructive sleep apnea and psychomotor vigilance task performance. *Nat. Sci. Sleep*, 2014, 6: 65–71.
- Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R. and Kupfer, D. J. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *J. Psychiatr. Res.*, 1989, 28: 193–213.
- Chervin, R. D. and Aldrich, M. S. The Epworth Sleepiness Scale may not reflect objective measures of sleepiness or sleep apnea. *Neurology*, 1999, 52: 125–131.
- Dinges, D. F. and Powell, J. W. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behav. Res. Methods Instrum. Comput.*, 1985, 17: 652–655.
- Dinges, D. F., Pack, F., Williams, K. *et al.* Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4–5 hours per night. *Sleep*, 1997, 20: 267–277.
- Doran, S. M., Van Dongen, H. P. A. and Dinges, D. F. Sustained attention performance during sleep deprivation: evidence of state instability. *Arch. Ital. Biol.*, 2001, 139: 253–267.
- Durmer, J. S. and Dinges, D. F. Neurocognitive consequences of sleep deprivation. *Sem. Neurol.*, 2005, 25: 117–129.
- Ftouni, S., Sletten, T. L., Howard, M. *et al.* Objective and subjective measures of sleepiness, and their associations with on-road driving events in shift workers. *J. Sleep Res.*, 2013, 22: 58–69.
- Grant, D. A. and Van Dongen, H. P. A. Individual differences in sleep duration and responses to sleep loss. In: P. Shaw, M. Tafti and M. J. Thorpy (Eds) *The Genetic Basis of Sleep and Sleep Disorders*. Cambridge University Press, New York, 2013: 189–196.
- Horne, J. A. and Östberg, O. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *Int. J. Chronobiol.*, 1976, 4: 97–110.
- Jewett, M. E., Dijk, D. J., Kronauer, R. E. and Dinges, D. F. Dose-response relationship between sleep duration and human psychomotor vigilance and subjective alertness. *Sleep*, 1999, 22: 171–179.
- Johns, M. W. A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. *Sleep*, 1991, 14: 540–545.
- Johns, M. W. Reliability and factor analysis of the Epworth Sleepiness Scale. *Sleep*, 1992, 15: 376–381.
- Johns, M. W. Sleepiness in different situations measured by the Epworth Sleepiness Scale. *Sleep*, 1994, 17: 703–710.
- Kribbs, N. B. and Dinges, D. F. Vigilance decrement and sleepiness. In: J. R. Harsh and R. D. Ogilvie (Eds) *Sleep Onset Mechanisms*. American Psychological Association, Washington, DC, 1994: 113–125.
- Kuna, S. T., Maislin, G., Pack, F. M. *et al.* Heritability of performance deficit accumulation during acute sleep deprivation in twins. *Sleep*, 2012, 35: 1223–1233.
- Loh, S., Lamond, N., Dorrian, J., Roach, G. and Dawson, D. The validity of psychomotor vigilance tasks of less than 10-minute duration. *Behav. Res. Methods Instrum. Comput.*, 2004, 36: 339–346.
- Littner, M., Kushida, C. A., McDowell Anderson, W., Bailey, D., Berry, R. B., Davila, D. G., Johnson, S. F. (2003): Standards of Practice Committee of the American Academy of Sleep Medicine. Practice parameters for the role of actigraphy in the study of sleep and circadian rhythms: An update for 2002. *Sleep*, 26(3), 337–341.
- Miller, N. L., Matsangas, P. and Shattuck, L. G. Fatigue and its effect on performance in military environments. In: P. A. Hancock and J. L. Szalma (Eds) *Performance Under Stress*. Ashgate Publishing, Burlington, VT, 2008: 231–250.
- Miller, N. L., Shattuck, L. G. and Matsangas, P. Sleep and fatigue issues in continuous operations: a survey of U.S. Army Officers. *Behav. Sleep Med.*, 2011, 9: 53–65.
- Miller, N. L., Matsangas, P. and Kenney, A. The role of sleep in the Military: implications for training and operational effectiveness. In: J. H. Laurence and M. D. Matthews (Eds) *The Oxford Handbook of Military Psychology*. Oxford University Press, New York, 2012: 262–281.
- Olson, L., Cole, M. and Ambrogetti, A. Correlations among Epworth Sleepiness Scale scores, multiple sleep latency tests and psychological symptoms. *J. Sleep Res.*, 1998, 7: 248–253.
- Parks, P. D., Durand, G., Tsismenakis, A. J., Vela-Bueno, A. and Kales, S. N. Screening for obstructive sleep apnea during commercial driver medical examinations. *J. Occup. Environ. Med.*, 2009, 51: 275–282.
- Rosekind, M. R., Graeber, R. C., Dinges, D. F., Connel, L. J., Rountree, M. S. and Spinweber, C. L. *Crew Factors in Flight Operations IX: Effects of Planned Cockpit Rest on Crew Performance and Alertness in Long-Haul Operations*. NASA Ames Research Center, Moffett Field, CA, 1994.
- Rupp, T. L., Wesensten, N. J. and Balkin, T. J. Trait-like vulnerability to total and partial sleep loss. *Sleep*, 2012, 35: 1163–1172.
- Talmage, J. B., Hudson, T. B., Hegmann, K. T. and Thiese, M. S. Consensus criteria for screening commercial drivers for obstructive sleep apnea: evidence of efficacy. *J. Occup. Environ. Med.*, 2008, 50: 324–329.
- Van Dongen, H. P. A. Connecting the dots: from trait vulnerability during total sleep deprivation to individual differences in cumulative impairment during sustained sleep restriction. *Sleep*, 2012, 35: 1031–1033.
- Wyatt, J. K., Dijk, D. J., Ronda, J. M. *et al.* Interaction of circadian- and sleep/wake homeostatic-processes modulate psychomotor vigilance test (PVT) performance (Abstract). *Sleep Res.*, 1997, 26: 759.