

3-minute smartphone-based and tablet-based psychomotor vigilance tests for the assessment of reduced alertness due to sleep deprivation

Devon A. Grant^{1,2} · Kimberly A. Honn¹ · Matthew E. Layton^{1,3} · Samantha M. Riedy¹ · Hans P. A. Van Dongen¹

Published online: 20 June 2016
© Psychonomic Society, Inc. 2016

Abstract The psychomotor vigilance test (PVT) is widely used to measure reduced alertness due to sleep loss. Here, two newly developed, 3-min versions of the psychomotor vigilance test, one smartphone-based and the other tablet-based, were validated against a conventional 10-min laptop-based PVT. Sixteen healthy participants (ages 22–40; seven males, nine females) completed a laboratory study, which included a practice and a baseline day, a 38-h total sleep deprivation (TSD) period, and a recovery day, during which they performed the three different versions of the PVT every 3 h. For each version of the PVT, the number of lapses, mean response time (RT), and number of false starts showed statistically significant changes across the sleep deprivation and recovery days. The number of lapses on the laptop was significantly correlated with the numbers of lapses on the smartphone and tablet. The mean RTs were generally faster on the smartphone and tablet than on the laptop. All three versions of the PVT exhibited a time-on-task effect in RTs, modulated by time awake and time of day. False starts were relatively rare on all three PVTs. For the number of lapses, the effect sizes across 38 h of TSD were large for the laptop PVT

and medium for the smartphone and tablet PVTs. These results indicate that the 3-min smartphone and tablet PVTs are valid instruments for measuring reduced alertness due to sleep deprivation and restored alertness following recovery sleep. The results also indicate that the loss of sensitivity on the 3-min PVTs may be mitigated by modifying the threshold defining lapses.

Keywords Sleep deprivation · Psychomotor vigilance · Mobile PVT

Electronic supplementary material The online version of this article (doi:10.3758/s13428-016-0763-8) contains supplementary material, which is available to authorized users.

✉ Devon A. Grant
devongrant@wsu.edu

¹ Sleep and Performance Research Center and Elson S. Floyd College of Medicine, Washington State University, Spokane, WA, USA

² Sleep and Performance Research Center, Washington State University, Spokane, WA, USA

³ Program of Excellence in Addictions Research, Washington State University, Spokane, WA, USA

The psychomotor vigilance test (PVT; Dinges & Powell, 1985) is widely used to measure reduced alertness due to sleep loss (Dorrian, Rogers, & Dinges, 2005). The PVT is a simple reaction time task with a random inter-trial interval that, conventionally, ranges between 2 and 10 s. Traditionally, the task is administered on a custom-made device or a desktop or laptop personal computer, with a fixed task duration of 10 min per test session. The baseline response time (RT) distribution on the PVT has a long right-hand tail, and both sleep deprivation and time on task are associated with further skewing of the RT distribution to the right (Dinges & Kribbs, 1991; Doran, Van Dongen, & Dinges, 2001; Ratcliff & Van Dongen, 2011). For sleep deprivation, this effect is well captured by counting the number of lapses of attention, conventionally defined as RTs greater than 500 ms (Doran et al., 2001; Dorrian et al., 2005). The number of lapses on the PVT per test session is a sensitive measure of instability in sustained attention due to sleep deprivation (Balkin et al., 2004; Lim & Dinges, 2008; Van Dongen, Belenky, & Krueger, 2011a; Van Dongen, Maislin, Mullington, & Dinges, 2003). Also, the impact of sleep deprivation on PVT lapses shows large, trait-like individual differences (Van Dongen, Baynard, Maislin, & Dinges, 2004).

Although the 10-min personal computer version of the PVT is routinely used in laboratory studies, its use in field studies has been hindered by hardware and time requirements. Shorter, 5-min versions of the PVT administered on hand-held devices have been shown to provide results that are comparable to the standard 10-min version (Lamond, Dawson, & Roach, 2005; Honn, Riedy, & Grant, 2015; Thorne et al., 2005) although yielding smaller effect sizes. With relatively minor adjustments of the technical specifications of the PVT, further reductions in task duration have been shown to be possible without loss of functional validity (Basner & Dinges, 2012; Basner, Mollicone, & Dinges, 2011; Basner & Rubinstein, 2011). Basner et al. (2011) developed the brief PVT (PVT-B), which shortened the test duration to 3 min. To mitigate the reduced sensitivity with the shorter task duration, they decreased the intertrial interval from 2–10 s to 2–5 s. Furthermore, they reduced the lapse threshold from 500 to 355 ms.

On the basis of this work, two new, portable, 3-min versions of the PVT were implemented: one on a smartphone and one on a tablet. Here, we validated the 3-min smartphone-based and tablet-based PVTs against a traditional 10-min laptop-based PVT. To be able to investigate a wide range of degrees of sleep loss efficiently, we performed the validation on a laboratory-controlled total sleep deprivation (TSD) study.

Method

Participants

Sixteen healthy participants (nine females, seven males) completed the study. Participants' ages ranged from 22 to 40 years ($M = 29.0$ years, $SD = 6.2$). One participant reported to be left-handed and two participants reported to be ambidextrous; the others reported to be right-handed.

Participants were screened by telephone interview and during two laboratory screening sessions to ensure they met eligibility criteria. Participants were physically and psychologically healthy and free of drugs and alcohol, as assessed by physical examination, blood chemistry, urinalysis, breathalyzer, history, and questionnaires. They had no clinically relevant history of psychiatric illness and no clinically relevant history of brain injury, and reported no learning disability. Participants were not vision impaired (or were corrected to normal). They were not current tobacco users and reported no current medical or drug treatment (excluding oral contraceptives). They had no history of drug or alcohol abuse in the past year and no history of methamphetamine abuse. Female participants were not pregnant.

Participants reported to be good sleepers, with habitual sleep durations between 6 and 10 h per day and regular wake-up times between 06:00 and 09:00. They had no sleep

or circadian disorders as assessed by questionnaires and history. They had not traveled across time zones within 1 month of entering the study and reported no shift work within three months of entering the study.

Each day during the seven days prior to admission into the laboratory portion of the study, participants completed a sleep/wake diary and called a time-stamped voicemail box to report their bed and wake times before going to bed in the evening and upon awakening in the morning. They also continually wore an Actiwatch-2 wrist actigraph (Respironics, Bend, OR) to measure their rest/activity patterns. Participants' mean daily sleep duration in the week before the laboratory study as estimated by the wrist actigraph was 7.3 h ($SD = 1.1$ h). Participants' average wake-up time as estimated by actigraphy was 07:32 ($SD = 0.9$ h).

Participants were financially compensated for their study participation. All participants gave written, informed consent, and the study was approved by the Institutional Review Board of Washington State University.

Experimental design

The study took place inside the Human Sleep and Cognition Laboratory of the Sleep and Performance Research Center at Washington State University Spokane. Participants participated in groups of up to four and were assigned their own room for performance testing and for scheduled sleep periods. They were in the laboratory for five days (four nights) for a total of 96 h consecutively. The laboratory environment was carefully controlled, with fixed ambient temperature ($M = 22$ °C, $SD = 1$ °C), and fixed, dim light levels (<100 lux) during periods of scheduled wakefulness. Participants were not allowed to use phones, computers, or otherwise communicate with the outside world, and no visitors were allowed during the study.

Participants reported to the laboratory on a designated day at 14:00. Admission procedures included urine and breathalyzer testing to verify that participants were free of drugs and alcohol. The at-home actigraphy was reviewed to verify consistent sleep/wake patterns in the week before the study. Participants were briefed on the study procedures, and were instructed to give their best effort on all performance tests.

Figure 1 shows a schematic of the study design. The first day in the laboratory served as a practice day. Participants went to bed at 22:00, with scheduled awakening at 08:00 the following day. The second day served as a baseline day. Participants went to bed at 22:00 for baseline sleep, with scheduled awakening at 08:00. Beginning on the third day, participants underwent 38 h of TSD. At the end of the 38-h TSD period, on the fourth day, participants went to bed at 22:00 for recovery sleep, with scheduled awakening at 08:00. Participants were discharged from the laboratory on the fifth day at 14:00.

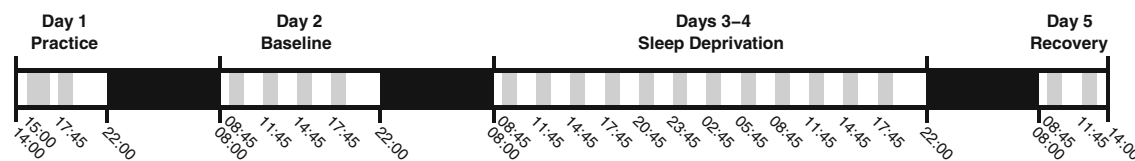


Fig. 1 Schematic of the five-day (four-night) laboratory protocol. Participants entered the laboratory at 14:00 on Day 1 and remained until 14:00 on Day 5. The distinct phases of the study are indicated above the schematic. The tick marks below the schematic indicate study

start and end times (14:00) as well as bedtimes (22:00) and wake-up times (08:00). All other clock times indicate the start times of the test blocks. Black, scheduled sleep; gray, performance testing

Participants were tested on a variety of neurobehavioral performance tests and subjective scales at regular intervals during scheduled wakefulness (see Fig. 1). Between performance test blocks, participants engaged in nonvigorous activities in the laboratory (watching movies, playing board games, or talking with other participants and research staff). Meals were served every 4 h of scheduled wakefulness; caffeine was not permitted. Participants were allowed to shower at 14:00 on the third day of the study and again after discharge on the fifth day. They were behaviorally monitored by trained research assistants throughout the laboratory study. The research assistants also ensured that participants did not fall asleep during scheduled waking periods.

During the practice day, participants were trained on the 10-min laptop PVT under the supervision of a staff member. They were also trained on filling out a paper-and-pencil NASA Task Load Index (NASA-TLX) workload questionnaire (Hart & Staveland, 1988). Participants were then trained on the 3-min smartphone PVT, and subsequently filled out a smartphone-based version of the Karolinska Sleepiness Scale (KSS; Åkerstedt & Gillberg, 1990) and another paper-and-pencil NASA-TLX. After a short break, participants were trained on a set of ten neurobehavioral performance tests on the tablet, which began with the 3-min tablet PVT. After completion of the tablet tests, participants filled out the NASA-TLX and a paper-and-pencil Scale of Positive and Negative Experience (SPANE; Diener et al., 2009). After a break following the training session, participants repeated the battery of tests to practice. The data from the training and practice test blocks were not used for analyses.

During all subsequent periods of scheduled wakefulness, from the second day at 08:45 through to the fifth day at 11:45, participants performed the test battery every 3 h. Each test block included the 3-min smartphone PVT, followed by the KSS and a 10-min break; the 10-min laptop PVT, followed by a 20-min break; and the 3-min tablet PVT, followed by nine additional neurobehavioral performance tests (which took approximately 20 min to complete) and the SPANE. Participants completed the NASA-TLX after each administration of the PVT during the 08:45 and 14:45 test blocks.

Of primary interest were the data from the three PVT versions—specifically, those that were administered in the 12 test

blocks of the 38-h TSD period beginning on the third day, and in the two test blocks of the recovery period (see Fig. 1).

Testing setup and implementations

Participants performed all PVTs at a desk in their assigned room. For each test block, the smartphone, laptop, or tablet device was brought into the participant's bedroom and placed on the desk in a marked location. The laptop and tablet were centrally located on the desk. The smartphone was placed to the left or right, depending on the participant's handedness, such that the smartphone was on the side of the participant's dominant hand (for the ambidextrous participants, it was placed on the right).

The laptop displays and the smartphones and tablets were angled upward toward the participant, to ensure a comfortable and unobstructed view that was free of glare. The laptop displays were at an angle of approximately 65° relative to the surface of the desk. The smartphones were attached to a rubber stand, keeping them steady on the desk at an angle of approximately 45°. The tablets were put in leather cases with an angle lock position kickstand set at approximately 33°. Each participant's chair was adjusted such that the participant could comfortably operate each device. Participants were instructed to maintain a stable posture with their back against the chair, feet flat on the floor. Participants were monitored during testing with camera views of their rooms (wide angle) and of their faces (narrow angle) to verify compliance.

During PVT testing, participants responded to a stimulus in the form of a rolling millisecond counter, which appeared in yellow inside a red box on a black screen. The stimulus appeared at random inter-trial intervals (from previous response to next stimulus presentation) ranging from 2 to 10 s for the 10-min laptop PVT, and ranging from 2 to 5 s for the 3-min smartphone and tablet PVTs. Participants were instructed to respond to the appearance of the stimulus as quickly as they could while trying to avoid false starts (errors of commission). They responded by pressing the space bar (on the laptop) or finger-tapping the touch screen below (not over) the stimulus (on the smartphone and tablet), using the index finger of their dominant hand. A valid response stopped the rolling millisecond counter, which then displayed the RT for 1 s.

The laptops used in the study were Latitude model E6420 (Dell Inc., Round Rock, TX) with 14-in. displays, running the Windows 7 operating system (Microsoft Inc., Redmond, WA). The laptops were individually calibrated and had Cognition software version 2.0.5.9 (Pulsar Informatics, Inc., Philadelphia, PA) with the PVT programmed to published specifications of the 10-min PVT (Basner & Dinges, 2011).

The smartphones used in the study were Galaxy S3 model GT-I9300 (Samsung Electronics Co., Ltd., Suwon, South Korea), running the Android operating system version 4.1.2 (Google Inc., Mountain View, CA). The phone function was disabled and clock time was not visible during testing. The smartphones were individually calibrated and had PVT Mobile software version 1.0.15 (Pulsar Informatics, Inc., Philadelphia, PA) with the PVT programmed to published specifications of the 3-min PVT (Basner et al., 2011).

The tablets used in the study were iPad mini model A1432, running operating system version 6.1.3 (10B329; Apple Inc., Cupertino, CA). To avoid eyestrain in the dim light environment of the laboratory, the brightness on the tablet display was set at 70 % of maximum. The tablets were individually calibrated and had Joggle Research software, version 2.4 (Joggle Research, Inc., Seattle, WA), with the PVT programmed to published specifications of the 3-min PVT (Basner et al., 2011).

Statistical analyses

Data analyses focused on validation of the 3-min smartphone PVT and the 3-min tablet PVT against the 10-min laptop PVT across the fourteen 3-h test blocks in the 38-h TSD period and following recovery sleep (see Fig. 1). Due to equipment failure, the laptop-based PVT for one test and the tablet-based PVT data for two tests (all different participants) could not be used. All analysis procedures were robust to these three missing data points.

Sleep deprivation produces a systematic skewing of the RT distribution to the right, which has been shown to be dominated by a change in a single parameter of the underlying statistical distribution as a function of time awake (Ratcliff & Van Dongen, 2011). This effect is captured well by the number of lapses, defined as RTs greater than 500 ms for the laptop PVT (Dorrian et al., 2005), and as RTs greater than 355 ms for the smartphone and tablet PVTs (Basner et al., 2011). The number of lapses was analyzed as the primary outcome variable. Other global outcome variables analyzed were the means of the RTs, which were included to investigate general response time on each PVT version; and the numbers of false starts, defined as any response in the absence of a stimulus or within the first 100 ms of stimulus presentation, which was included because false starts are not part of the RT distribution and therefore not captured by the number of lapses or the mean of the RTs. The three global outcome

variables were first analyzed using mixed-effects analysis of variance (ANOVA; Van Dongen, Maislin, & Dinges, 2004) with a fixed effect for test blocks (1–14) and a random effect over participants on the intercept.

To validate the smartphone-based and tablet-based PVTs against the laptop PVT, the same laptop PVT data were re-analyzed without test blocks as fixed effect, but with the smartphone PVT or tablet PVT data on the same global outcome variable included as a covariate. The statistical significance of the covariate was determined, and the variance (within and between participants) explained by the covariate was assessed and statistically tested with a variance ratio test. As an additional validation analysis, the consistency of trait-like individual differences in global outcome variables between the different PVT versions was assessed. To this end, rank-order correlations were calculated over participants (averaged across test blocks during TSD and following recovery).

Additionally, the time-on-task effect was analyzed for each PVT version across all test blocks during TSD and following recovery, by dividing the task duration into 1-min bins. The raw RT data were analyzed using mixed-effects ANOVA with fixed effect for test blocks (1–14), 1-min bins (1–10 for the laptop PVT; 1–3 for the other PVTs), and their interaction, and a random effect over participants on the intercept.

The effects of TSD on the RT distributions were examined for the three PVT platforms by comparing data aggregated over daytime test blocks (08:45, 11:45, 14:45, and 17:45 combined) during the first day (well-rested) versus the second day (sleep-deprived) of the 38-h TSD period.

Finally, for the number of lapses, the effect size for the effect of test blocks specifically during the 38-h TSD period was determined in a mixed-effects ANOVA with a fixed effect for test blocks (1–12) and a random effect over participants on the intercept. The effect size was assessed using Cohen's f^2 (Selya, Rose, Dierker, Hedeker, & Mermelstein, 2012), where $f^2 \geq .15$ signifies moderate effect size and $f^2 \geq .35$ signifies large effect size (Cohen, 1988). To investigate whether using different thresholds for defining a lapse could increase the effect sizes on the different versions of the PVT, the thresholds were varied from 200 to 500 ms in steps of 5 ms, and Cohen's f^2 was reassessed.

Results

Number of lapses

Figure 2 (top) shows the data for PVT numbers of lapses across the 14 test blocks during TSD and after recovery. There were significant effects of test block for the laptop PVT [$F(13, 194) = 10.6, p < .001$], the smartphone PVT [$F(13, 195) = 5.2, p < .001$], and the tablet PVT [$F(13, 193) = 5.9, p < .001$]. The temporal profile of the number of lapses

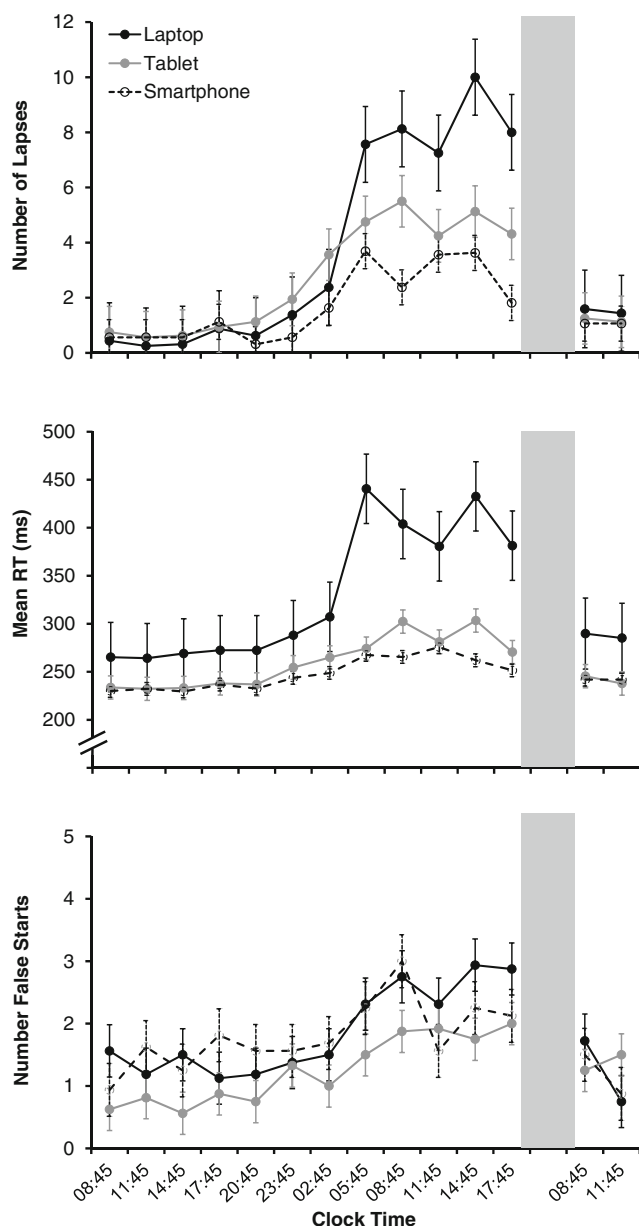


Fig. 2 Means \pm SEs for numbers of lapses (top), mean RTs (middle), and numbers of false starts (bottom) on the three different PVT platforms across the 38 h of TSD and following recovery sleep (indicated by the gray bar)

on each version of the PVT during the 38-h TSD period reflected the influence of time awake interacting with time of day (Van Dongen & Dinges, 2005). Due to the differences in task duration between the laptop PVT (10 min) and the smartphone and tablet PVTs (3 min), and despite the fact that a more conservative lapse threshold was used for the laptop (500 ms) than for the smartphone and tablet (355 ms), the number of lapses was considerably higher on the laptop PVT than on the smartphone and tablet PVTs during the second day of TSD. After recovery sleep, the number of lapses returned to near the levels seen at the beginning of the 38-h

TSD period for all three PVT versions (cf. Van Dongen & Dinges, 2005).

For the validation analysis of the smartphone PVT, lapses on the smartphone were found to be a significant covariate for lapses on the laptop [$F(1, 206) = 149.1, p < .001$]. The number of lapses on the smartphone explained 38.4 % of the variance in the number of lapses on the laptop [$F(1, 206) = 130.3, p < .001$]. That is, within and between subjects the correlation between the smartphone and laptop PVTs for number of lapses was 0.62. Furthermore, focusing specifically on individual differences, the rank-order correlation of PVT lapses over participants (averaged across test blocks) was found to be $\rho = 0.61$ ($p = .012$).

For the validation analysis of the tablet PVT, lapses on the tablet were found to be a significant covariate for lapses on the laptop [$F(1, 204) = 131.3, p < .001$]. The number of lapses on the tablet explained 40.7 % of the variance in the number of lapses on the laptop [$F(1, 204) = 141.5, p < .001$]. Thus, within and between subjects, the correlation between the tablet and laptop PVTs for number of lapses was 0.64. Furthermore, the rank-order correlation of PVT lapses over participants (averaged across test blocks) was found to be $\rho = 0.69$ ($p = .003$).

Mean response time

Figure 2 (middle) shows the data for PVT mean RTs across the 14 test blocks during TSD and after recovery. We found significant effects of test block for the laptop PVT [$F(13, 194) = 5.1, p < .001$], the smartphone PVT [$F(13, 195) = 10.0, p < .001$], and the tablet PVT [$F(13, 193) = 6.2, p < .001$]. As was the case for lapses, the temporal profile of mean RT on each version of the PVT during the 38-h TSD period reflected the influence of time awake interacting with time of day. Given the longer task duration of the laptop PVT (10 min) than of the smartphone and tablet PVTs (3 min), and due to the time-on-task effect interacting with time awake and time of day (as further discussed below), the mean RT was consistently higher on the laptop PVT than on the smartphone and tablet PVTs throughout the 38-h TSD period. After recovery sleep, mean RT returned to near the levels seen at the beginning of the 38-h TSD period for all three PVT versions.

Mean RT on the smartphone was a significant covariate for mean RT on the laptop [$F(1, 206) = 121.0, p < .001$]. Mean RT on the smartphone explained 32.5 % of the variance in mean RT on the laptop [$F(1, 206) = 100.7, p < .001$]. That is, within and between subjects the correlation between the smartphone and laptop PVTs for mean RT was .57. In addition, the rank-order correlation of mean RT between the smartphone and laptop PVTs over participants (averaged across test blocks) was $\rho = .76$ ($p < .001$).

Mean RT on the tablet was also a significant covariate for mean RT on the laptop [$F(1, 204) = 54.2, p < .001$]. Mean RT

on the tablet PVT explained 22.0 % of the variance in mean RT on the laptop PVT [$F(1, 204) = 58.9, p < .001$]. Thus, within and between subjects, the correlation between the tablet and laptop PVTs for mean RT was .47. In addition, the rank-order correlation of mean RT between the tablet and laptop PVTs over participants (averaged across test blocks) was $\rho = .81$ ($p < .001$).

Number of false starts

Figure 2 (bottom) shows the data for PVT number of false starts across the 14 test blocks during TSD and after recovery. We found significant effects of test block for the laptop PVT [$F(13, 194) = 3.7, p < .001$], the smartphone PVT [$F(13, 195) = 2.6, p = .002$], and the tablet PVT [$F(13, 193) = 2.8, p = .001$]. During the 38-h TSD period, false starts on the smartphone and tablet PVTs showed the same temporal profile as false starts on the laptop PVT, with a steady increase across time awake, but remaining relatively rare. After recovery sleep, false starts returned to near the levels seen at the beginning of the 38-h TSD period for all three PVT versions.

The number of false starts on the smartphone was a significant covariate for the number of false starts on the laptop [$F(1, 206) = 6.5, p = .012$]. However, false starts on the smartphone explained only 3.8 % of the variance in false starts on the laptop PVT [$F(1, 206) = 9.2, p = .003$]. Thus, within and between subjects, the correlation between the smartphone and laptop PVTs for false starts was .20. The rank-order correlation of false starts between the smartphone and laptop PVTs over participants (averaged across test blocks) was $\rho = .24$ ($p = .36$).

The number of false starts on the tablet was not a significant covariate for the number of false starts on the laptop [$F(1, 204) = 2.7, p = .104$]. False starts on the tablet explained only 1.4 % of the variance in false starts on the laptop PVT [$F(1, 204) = 4.0, p = .047$]. Thus, within and between subjects, the correlation between the tablet and laptop PVTs for false starts was .12. The rank-order correlation of false starts between the tablet and laptop PVTs over participants (averaged across test blocks) was $\rho = .23$ ($p = .40$).

Time-on-task effect

Figure 3 shows the results of the mean RT time-on-task analyses, in 1-min bins of the task duration of each PVT version across the 14 test blocks during TSD and after recovery. For the 10-min laptop PVT, we observed significant effects of test block [$F(13, 20541) = 67.1, p < .001$], bin [$F(9, 20541) = 21.4, p < .001$], and their interaction [$F(117, 20541) = 1.9, p < .001$]. Similarly, for the 3-min smartphone PVT, there were significant effects of test block [$F(13, 9975) = 43.1, p < .001$], bin [$F(2, 9975) = 38.7, p < .001$], and their interaction [$F(26, 9975) = 2.7, p < .001$]. For the 3-min tablet PVT, significant

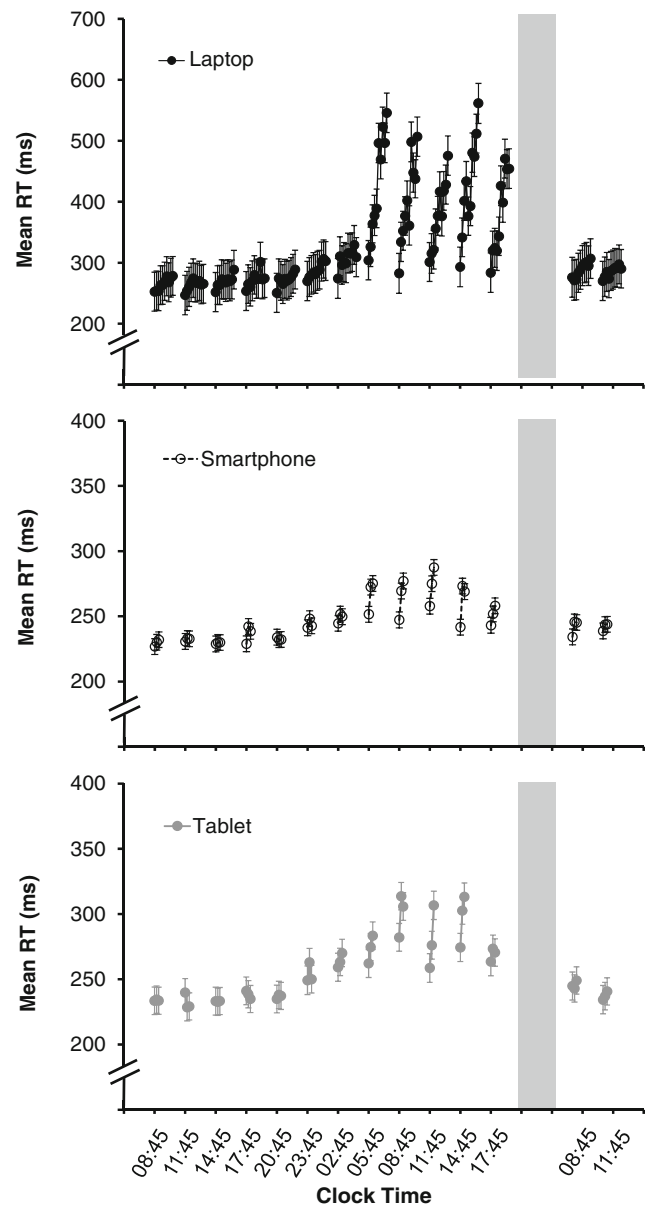


Fig. 3 Means \pm SEs for RTs in 1-min bins on the 10-min laptop PVT (top), the 3-min smartphone PVT (middle), and the 3-min tablet PVT (bottom). Data are plotted against the start times of the test blocks across the 38-h TSD period and following recovery (indicated by the gray bar). Placement of the 1-min bins in each test is not to scale on the clock time axis. Also note the difference in the vertical scales between the top graph and the other two graphs

effects of test block [$F(13, 9903) = 25.4, p < .001$] and bin [$F(2, 9903) = 5.9, p = .003$] emerged, but not for their interaction [$F(26, 9903) = 1.3, p = .17$].

All three PVT versions showed a general increase of mean RTs with progressing time on task, which was reset by the rest breaks between the tests and test blocks. For each PVT version, the time-on-task effect was amplified as a function of time awake and time of day (Van Dongen et al., 2011a; Wesensten, Belenky, Thorne, Kautz, & Balkin, 2004). For the 3-min table PVT, the interaction of the test blocks with

the 1-min bins did not reach statistical significance, but the time-on-task patterns nonetheless resembled those of the 3-min smartphone PVT (Fig. 3). Furthermore, for all three PVT versions, the changes in the time-on-task effect over test blocks mirrored the influences of time awake and time of day on global mean RT (cf. Fig. 2, middle). Following recovery sleep, the time-on-task effect was diminished, approaching the magnitude of the effect observed at the beginning of the 38-h TSD period.

RT distributions

Figure 4 shows the RT distributions on the three PVT platforms, comparing daytime test blocks (08:45, 11:45, 14:45, and 17:45 combined) during the first day (well-rested) to the second day (sleep-deprived) of the 38-h TSD period. Each PVT version demonstrated the typical increase in skewing of the RT distribution to the right after sleep deprivation (Dinges & Kribbs, 1991; Ratcliff & Van Dongen, 2011).

Effect sizes

The effect size for the change in the primary outcome variable, number of lapses, during the 38-h TSD period was assessed using Cohen's f^2 . We found that the effect size was large for the laptop PVT ($f^2 = .42$), medium for the smartphone PVT ($f^2 = .24$), and medium for the tablet PVT ($f^2 = .24$).

To determine whether the sensitivity of each of the three versions of the PVT could be increased, the threshold for defining lapses was varied from 200 to 500 ms, in steps of 5 ms, and the assessment of effect size was repeated. Figure 5 shows how Cohen's f^2 varied for each PVT version as a function of the lapse threshold. For the 10-min laptop PVT, the effect size was maximized at a lapse threshold of 390 ms ($f^2 = .64$). For the 3-min smartphone PVT, the effect size was maximized at a lapse threshold of 220 ms ($f^2 = .33$). For the 3-min tablet PVT, the effect size was maximized at a lapse threshold of 280 ms ($f^2 = .37$).

Discussion

The purpose of the present laboratory study was to validate two 3-min versions of a PVT, one implemented on a smartphone and the other on a tablet, against a gold-standard 10-min PVT version on a laptop. Across a 38-h period of TSD, as well as following recovery sleep, the three PVT devices demonstrated parallel profiles of sleep-deprivation-induced performance impairment, with increases in numbers of lapses, mean RTs, and false starts (Fig. 2). The temporal profiles of the number of lapses and mean RT on each version of the PVT reflected the well-established influence of time awake interacting with time of day (Daan, Beersma, &

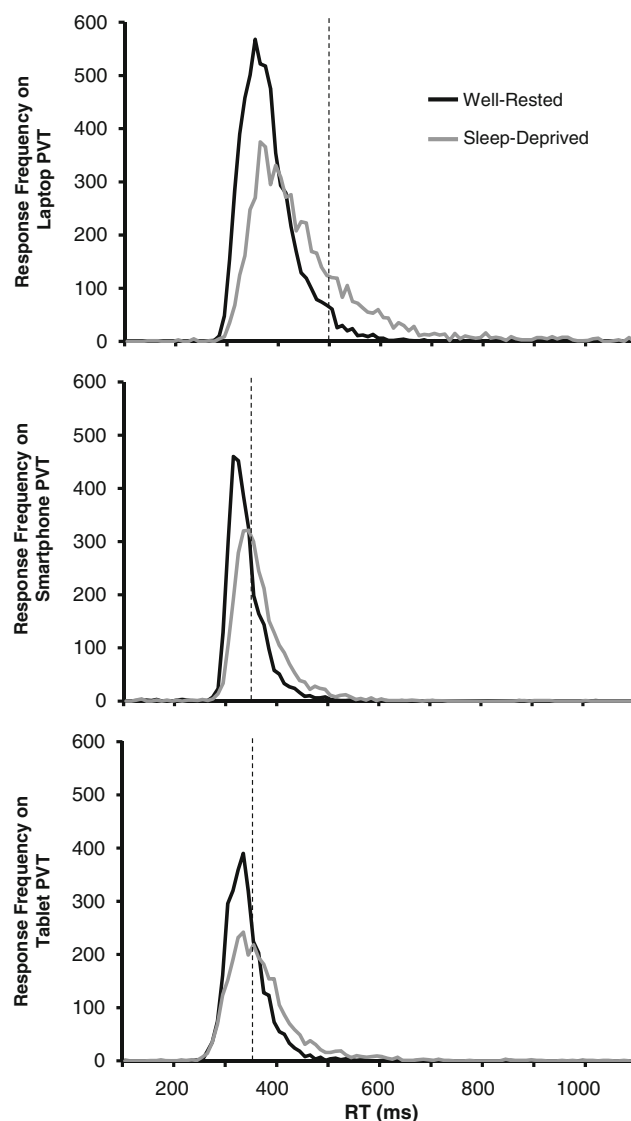


Fig. 4 RT distributions during daytime test blocks of the first and second days of TSD, on the three different PVT platforms. Each graph compares the RT data pooled over participants and across the daytime test blocks (08:45, 11:45, 14:45, and 17:45) during the second day of the TSD period (sleep-deprived; gray) to the RT data collapsed over the equivalent daytime test blocks 24 h earlier (well-rested; black). Data are plotted in 10-ms bins. The dashed vertical lines indicate the lapse thresholds used in the primary analyses for each of the PVT versions

Borbély, 1984; Dijk, Duffy, & Czeisler, 1992). As expected, recovery sleep restored performance on each version of the PVT to near baseline levels (Van Dongen & Dinges, 2005).

Because of the shorter task duration of the 3-min PVTs, less time was available for the time-on-task effect to be expressed than in the 10-min PVT, resulting in shorter mean RTs on the 3-min PVTs overall (Fig. 2, middle). (For reference, a head-to-head comparison of the 3-min PVTs with the data from the first 3 min of the 10-min PVT can be found in the [supplemental material](#).) All three PVT versions showed the same general dynamic in the time-on-task effect (Fig. 3), with stereotypical

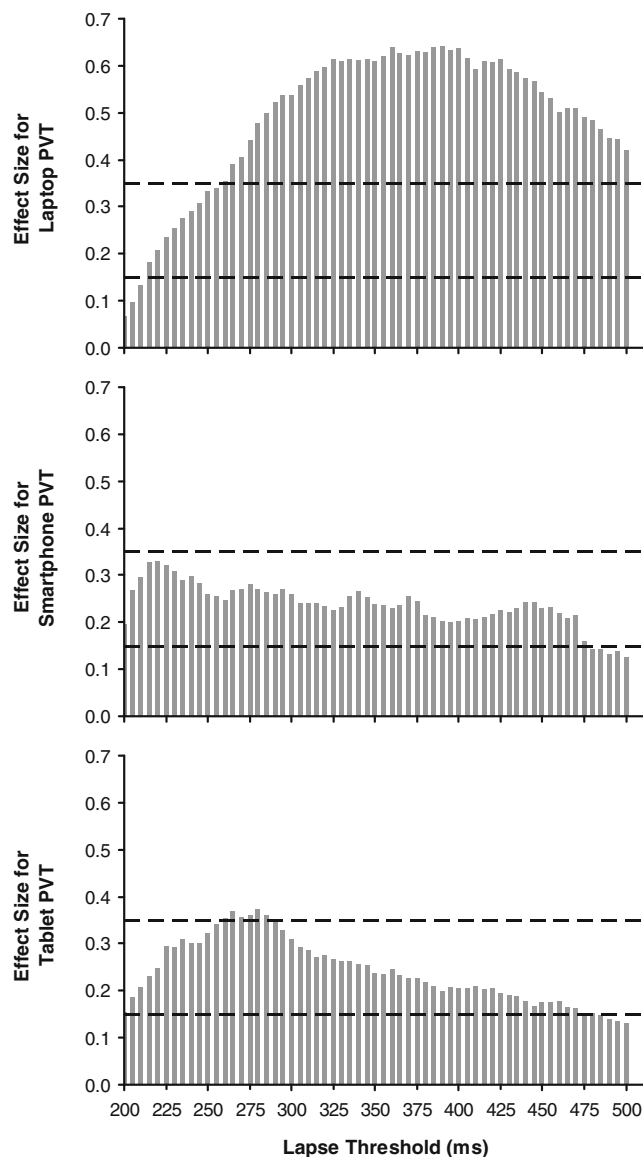


Fig. 5 Cohen's f^2 effect sizes for the effect of test block on the numbers of lapses, as a function of the lapse threshold, on the three different PVT platforms. Dashed lines delineate effect sizes in the medium ($.15 \leq f^2 < .35$) and large ($f^2 \geq .35$) ranges

amplification of the effect as a function of time awake and time of day (Van Dongen et al., 2011a; Wesensten et al., 2004) and returning to baseline patterns after recovery sleep. Individual differences in performance were comparable across the three devices, such that trait vulnerability to reduced alertness due to sleep deprivation (Van Dongen, Baynard, et al., 2004) was captured by each of the three PVT versions.

Each version of the PVT demonstrated the typical increase in the skewing of the RT distribution to the right after sleep deprivation (Dinges & Kribbs, 1991; Ratcliff & Van Dongen, 2011). This was less pronounced in the 3-min PVTs than in the 10-min PVT (Fig. 4), in part because the 3-min PVTs presented fewer stimuli and consequently exhibited lower signal-to-noise ratio. In the design of the 3-min PVTs, this problem was

already mitigated somewhat by increasing the stimulus density through reduction of the inter-trial interval (Basner et al., 2011). Not only does this increase the number of data points that make up the RT distribution, it also enhances the impact of sleep loss by promoting activity-driven (i.e., use-dependent) performance impairment (Van Dongen, Belenky, & Krueger, 2011b). Even so, the shorter task duration of the 3-min PVTs limited the magnitude of the time-on-task effect (Fig. 3), which made the reduction of the inter-trial interval only partially effective in maintaining sensitivity to sleep deprivation.

Focusing on the number of lapses as the primary outcome variable, further improvement in sensitivity to TSD was achieved by reducing the lapse threshold from 500 ms for the 10-min PVT to 355 ms for the 3-min PVTs (Basner et al., 2011), so that a greater number of RTs fell into the lapse domain of the RT distribution. This helped magnify the difference in the number of lapses after sleep deprivation relative to the well-rested state (Fig. 4). We showed that across 38 h of TSD, the effect size associated with this difference could be improved even more (Fig. 5) by selecting lapse thresholds of 220 ms for the 3-min smartphone PVT and 280 ms for the 3-min tablet PVT. Improvement could also be achieved for the 10-min laptop PVT, by setting the lapse threshold to 390 ms.

It should be noted that the study was conducted with healthy young adults as participants. Whether the specific lapse thresholds suggested here are optimal in other populations—such as older adults, who tend to have reduced speed of cognitive processing (Salthouse, 2000)—remains to be investigated. Also, the order in which the three versions of the PVT were administered within each test block was fixed: first the 3-min smartphone PVT, followed by a 10-min break; then the 10-min laptop PVT, followed by a 20-min break; and then the 3-min tablet PVT. Although each break was at least twice as long as the PVT it followed, it is possible that some carry-over effects occurred within each test block. This may have contributed to the somewhat greater magnitude of TSD effects on the 3-min tablet PVT than on the 3-min smartphone PVT (Fig. 2), in addition to the different device characteristics. Carryover effects within test blocks may also have influenced the optimal lapse threshold of 280 ms that we found for the tablet PVT. Judging by the systematic resetting of the time-on-task effect at the beginning of each test block (Fig. 3), however, there seems to be no evidence of any substantive carry-over effects from one test block to the next—so we have no reason to believe that the temporal profiles of the global outcome variables were affected.

In our laboratory-based study, we found that although the sensitivity of the 3-min PVTs is not as high as that of the standard 10-min PVT, the 3-min smartphone and tablet PVTs are reliable for measuring the decrease of alertness due to sleep deprivation and the restoration of alertness following recovery sleep. Although we used a TSD paradigm to

investigate PVT outcomes across a wide range of degrees of sleep loss, we expect that our findings generalize to conditions of chronic partial sleep restriction, which are more commonly encountered in the real world. It has been shown experimentally (Belenky et al., 2003; Van Dongen et al., 2003) and mathematically (McCauley et al., 2009) that the effects of TSD and chronic sleep restriction on PVT performance share the same underlying dynamics.

To what extent our laboratory findings may extend to field settings requires additional experimentation. Previous studies (Lamond, Dawson, & Roach, 2005; Thorne et al., 2005; Van Dongen & Mollicone, 2013) demonstrated the utility of hand-held versions of the PVT in the field. However, the devices investigated in the present study were mounted on a stationary desk, which is not necessarily representative of how the PVT is administered in field settings. The use of a touch screen to register responses to stimuli increases the potential for accidental false starts, as has been found previously for a 5-min PVT implemented on a personal digital assistant (Honn et al., 2015). Although this was not found to be an issue in the present study (Fig. 2, bottom), nor in a recent field study of truck drivers who performed the 3-min smartphone PVT when they were not driving (Van Dongen & Mollicone, 2013), false starts could still be artificially increased in moving environments (e.g., commercial aviation). False starts have been interpreted as an index of effort to perform (Doran et al., 2001). In field studies employing the PVT, caution may be warranted with regard to that interpretation.

In conclusion, our results indicate that the 3-min smartphone PVT and the 3-min tablet PVT (Basner et al., 2011) are valid instruments for measuring reduced alertness due to sleep deprivation. Owing to their short task duration and implementation on hand-held devices that are also already used in real-world settings for other purposes, these versions of the PVT are likely to be useful and convenient as objective assays of performance instability in support of operational fatigue risk management (Satterfield & Van Dongen, 2013). Further improvement in the sensitivity to sleep deprivation of these tests can be achieved by adjusting the threshold used to define lapses of attention, or by considering other outcome variables designed to maximize effect size (Basner, McGuire, Goel, Rao, & Dinges, 2015).

Author Note We are grateful to an anonymous reviewer for suggesting the additional analyses in the [supplemental material](#). We thank the staff of the Human Sleep and Cognition Laboratory in the Sleep and Performance Research Center at Washington State University for their assistance in data collection. The smartphone-based and tablet-based PVTs used in the study were on loan from Pulsar Informatics, Inc. This company was not involved in the data analyses or manuscript writing. This research was supported by Office of Naval Research Grant No. N00014-13-C-0063.

References

- Åkerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, 52, 29–37.
- Balkin, T. J., Bliese, P. D., Belenky, G., Sing, H., Thorne, D. R., Thomas, M., ... Wesensten, N. J. (2004). Comparative utility of instruments for monitoring sleepiness-related performance decrements in the operational environment. *Journal of Sleep Research*, 13, 219–227.
- Basner, M., & Dinges, D. F. (2011). Maximizing sensitivity of the Psychomotor Vigilance Test (PVT) to sleep loss. *Sleep*, 34, 581–591.
- Basner, M., & Dinges, D. F. (2012). An adaptive-duration version of the PVT accurately tracks changes in psychomotor vigilance induced by sleep restriction. *Sleep*, 35, 193–202.
- Basner, M., McGuire, S., Goel, N., Rao, H., & Dinges, D. F. (2015). A new likelihood ratio metric for the psychomotor vigilance test and its sensitivity to sleep loss. *Journal of Sleep Research*.
- Basner, M., Mollicone, D., & Dinges, D. F. (2011). Validity and sensitivity of a brief psychomotor vigilance test (PVT-B) to total and partial sleep deprivation. *Acta Astronautica*, 69, 949–959.
- Basner, M., & Rubinstein, J. (2011). Fitness for duty: A 3 minute version of the psychomotor vigilance test predicts fatigue related to declines in luggage screening performance. *Journal of Occupation and Environmental Medicine*, 53, 1146–1154.
- Belenky, G., Wesensten, N. J., Thorne, D. R., Thomas, M. L., Sing, H. C., Redmond, D. P., ... Balkin, T. J. (2003). Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: A sleep dose-response study. *Journal of Sleep Research*, 12, 1–12.
- Cohen, J. E. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Erlbaum.
- Daan, S., Beersma, D. G., & Borbély, A. A. (1984). Timing of human sleep: Recovery process gated by a circadian pacemaker. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology*, 246, R161–R183.
- Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D., & Diswas-Diener, R. (2009). New measures of well-being: Flourishing and positive and negative feelings. *Social Indicators Research*, 39, 247–266.
- Dijk, D. J., Duffy, J. F., & Czeisler, C. A. (1992). Circadian and sleep/wake dependent aspects of subjective alertness and cognitive performance. *Journal of Sleep Research*, 1, 112–117.
- Dinges, D. F., & Kribbs, N. B. (1991). Performing while sleepy: Effects of experimentally-induced sleepiness. In T. H. Monk (Ed.), *Sleep, sleepiness and performance* (pp. 97–128). New York: Wiley.
- 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, & Computers*, 17, 652–655.
- Doran, S. M., Van Dongen, H. P. A., & Dinges, D. F. (2001). Sustained attention performance during sleep deprivation: Evidence of state instability. *Archives Italiennes de Biologie*, 139, 253–267.
- Dorrian, J., Rogers, N. L., & Dinges, D. F. (2005). Psychomotor vigilance performance: Neurocognitive assay sensitive to sleep loss. In C. A. Kushida (Ed.), *Sleep deprivation: Clinical issues, pharmacology, and sleep loss effects* (pp. 39–70). New York: Marcel Dekker.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Advances in Psychology*, 52, 139–183.
- Honn, K. A., Riedy, S. M., & Grant, D. A. (2015). Validation of a portable, touch-screen psychomotor vigilance test. *Aerospace Medicine and Human Performance*, 86, 428–434.
- Lamond, N., Dawson, D., & Roach, G. D. (2005). Fatigue assessment in the field: Validation of a hand-held electronic psychomotor vigilance task. *Aviation, Space, and Environmental Medicine*, 76, 486–489.
- Lim, J., & Dinges, D. F. (2008). Sleep deprivation and vigilant attention. *Annals of the New York Academy of Sciences*, 1129, 305–322.
- McCauley, P., Kalachev, L. V., Smith, A. D., Belenky, G., Dinges, D. F., & Van Dongen, H. P. A. (2009). A new mathematical model for the

- homeostatic effects of sleep loss on neurobehavioral performance. *Journal of Theoretical Biology*, 256, 227–239.
- Ratcliff, R., & Van Dongen, H. P. A. (2011). Diffusion model for one-choice reaction-time tasks and the cognitive effects of sleep deprivation. *Proceedings of the National Academy of Sciences*, 108, 11285–11290.
- Salthouse, T. A. (2000). Aging and measures of processing speed. *Biological Psychology*, 54, 35–54.
- Satterfield, B. C., & Van Dongen, H. P. A. (2013). Occupational fatigue, underlying sleep and circadian mechanisms, and approaches to fatigue risk management. *Fatigue: Biomedicine, Health & Behavior*, 1, 118–136.
- Selya, A. S., Rose, J. S., Dierker, L. C., Hedeker, D., & Mermelstein, R. J. (2012). A practical guide to calculating Cohen's f^2 , a measure of local effect size, from PROC MIXED. *Frontiers in Psychology*, 3, 111.
- Thorne, D. R., Johnson, D. E., Redmond, D. P., Sing, H. C., Belenky, G., & Shapiro, J. M. (2005). The Walter Reed palm-held psychomotor vigilance test. *Behavior Research Methods*, 37, 111–118. doi:10.3758/BF03206404
- Van Dongen, H. P. A., Baynard, M. D., Maislin, G., & Dinges, D. F. (2004). Systematic interindividual differences in neurobehavioral impairment from sleep loss: Evidence of trait-like differential vulnerability. *Sleep*, 27, 423–433.
- Van Dongen, H. P. A., Belenky, G., & Krueger, J. M. (2011a). Investigating the temporal dynamics and underlying mechanisms of cognitive fatigue. In P. L. Ackerman (Ed.), *Cognitive fatigue: Multidisciplinary perspective on current research and future applications* (pp. 127–147). Washington, DC: American Psychological Association.
- Van Dongen, H. P. A., Belenky, G., & Krueger, J. M. (2011b). A local, bottom-up perspective on sleep deprivation and neurobehavioral performance. *Current Topics in Medicinal Chemistry*, 11, 2414–2422.
- Van Dongen, H. P. A., & Dinges, D. F. (2005). Sleep, circadian rhythms, and psychomotor vigilance. *Clinics in Sports Medicine*, 24, 237–249.
- Van Dongen, H. P. A., Maislin, G., Mullington, J. M., & Dinges, D. F. (2003). The cumulative cost of additional wakefulness: Dose–response effects of neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep*, 26, 117–129.
- Van Dongen, H. P. A., Maislin, G., & Dinges, D. F. (2004). Dealing with inter-individual differences in the temporal dynamics of fatigue and performance: Importance and techniques. *Aviation, Space, and Environmental Medicine*, 75, A147–A154.
- Van Dongen, H. P. A., & Mollicone, D. (2013). *Field study on the efficacy of the new restart provision for Hours of Service (Report No. RRR-13-058)*. Washington, DC: Federal Motor Carrier Safety Administration.
- Wesensten, N. J., Belenky, G., Thorne, D. R., Kautz, M. A., & Balkin, T. J. (2004). Modafinil vs. caffeine: Effects on fatigue during sleep deprivation. *Aviation, Space, and Environmental Medicine*, 75, 520–525.