Validation of a touchscreen psychomotor vigilance task

Lucia Arsintescu[a,c,*], Kenji H. Kato[b,c], Patrick F. Cravalho[a,c], Nathan H. Feick[a,c], Leland S. Stone[c], Erin E. Flynn-Evans[c]

[a] San Jose State University Research Foundation, 210 N. 4th Street, Fourth Floor San José, CA 95112, USA
[b] ASRC Research and Technology Solutions, LLC, 7000 Muirkirk Meadows Drive, Beltsville, MD 20705, USA
[c] NASA Ames Research Center, Moffett Blvd, Mountain View, CA 94035 CA, USA

ARTICLE INFO

ABSTRACT

Objective: The purpose of this study was to compare a psychomotor vigilance task developed for use on touchscreen devices with the original PVT-192 in conditions of acute sleep loss and circadian desynchronization.
Background: The Psychomotor Vigilance Task (PVT) is considered the gold standard fatigue detection test and is used frequently in fatigue research. With the rapid development of new technologies it is essential to develop a PVT available on different platforms such as touchscreen devices. The advantage of such PVT is that it can be implemented on small devices and can be easily used in field studies.
Methods: Ten participants completed a 5-min PVT (NASA-PVT) on a touchscreen device and a 5-min PVT on the original PVT-192. On the day of the experiment, participants arrived in the lab approximately two hours after their habitual wake time. Participants completed a constant routine protocol under dim lighting, while maintaining a constant posture. The 5-min PVT-192 and NASA-PVT were taken every two hours for at least 24 h.
Results: The NASA-PVT and PVT-192 were sensitive to extended wakefulness in the same manner. The reaction times were slower and the lapses were higher as time progressed on both NASA-PVT and PVT-192 ($p < 0.001$). Overall, there was a sharp decline in performance after 16 h of being awake which coincided with the time the participants were usually going to bed and the worst performance occurred after 24 h of wakefulness for both PVTs ($p < 0.001$).
Conclusions: Overall, our data suggest that the NASA-PVT is a valid tool for assessing fatigue in field studies.

1. Introduction

The Psychomotor Vigilance Task (PVT) is a reaction time test for measuring performance due to fatigue in laboratory and field studies (Lee et al., 2010; Van Dongen et al., 2003; Lamond et al., 2005). The first PVT, named PVT-192 (Ambulatory Monitoring, Ardsley, NY; Dinges and Powel, 1985) is still widely used in laboratory studies. The PVT-192 is a handheld test that involves pressing a button in response to a stimulus counter that appears on a small LCD screen. After a participant responds, the stimulus counter displays the reaction time (RT) in milliseconds for a one-second period. The inter-stimulus interval varies randomly from 2 to 10 s. The traditional task duration is 10 min, but tests of 20 min or longer have been used based on the design of the study (Veksler and Gunzelmann, 2017). Participants are instructed to press the button as soon as each stimulus is seen, in order to keep the RT number as low as possible, but not to press the button before stimulus presentation (which yields a false start [FS] warning on the display). If

the participant presses the wrong button to respond to the stimuli an error [ERR] message will appear on the display.

With the fast progress of new mobile technologies, several versions of the PVT have been developed for handheld devices (Thorne et al., 2005; Kay et al., 2013; Honn et al., 2015; Grant et al., 2017). The most commonly used handheld version of the PVT is a 5-min test that was developed for use on Palm OS (Palm Inc., Sunnyvale, CA) and validated against the PVT-192 (Thorne et al., 2005). The Palm-PVT has been validated in both laboratory and field studies investigating sleep deprivation and fatigue (Dorrian et al., 2006; Lamond et al., 2008; Kamimori et al., 2005; Lamond et al., 2005). The main advantage of the Palm-PVT was that it could be easily used in the field due to its small and easy to carry device. Although the Palm-PVT retained many features of PVT-192, there are differences between the two devices (Thorne et al., 2005). The most notable ones are that the stimulus on the Palm-PVT was designed to look as a target instead of a millisecond counter and the feedback is provided at the end of each session instead

of at the end of each trial. In addition, the Palm-PVT uses N discrete inter-stimulus intervals determined by a user-specified step size, which are randomized without replacement in blocks of 2N, while the PVT-192 uses a 2–10 s interval randomized based on a rectangular distribution. This yields a number of trials per minute that is determined more by the average fore- period (e.g., 6 s) than by variations in the average RT (e.g., 250–500 milliseconds; Thorne et al., 2005).

The outcomes from the PVT test that have been shown to be most sensitive to fatigue are mean reaction time, number of performance lapses (trials in which the participant failed to generate a response within 500 milliseconds (ms)), inverse reaction time (MEAN 1/RT), fastest 10% reaction time, and slowest 10% reaction time (Basner and Dinges, 2011). The PVT has no significant learning curve, making this task ideal for evaluating fatigue in field settings (Dinges and Kribbs, 1991).

Given that the Palm-PVT is not supported any longer and the increasing diversity of handheld devices, we developed a PVT (NASA-PVT) to be used on touchscreen devices that can be easily deployed in the field. The NASA-PVT was developed at NASA Ames Research Center on a fifth generation, 32-GB Apple iPod and has the same characteristics as the original PVT-192. In the present study we aimed to validate the 5-min NASA-PVT over 24 h of sleep deprivation.

2. Methods

2.1. Procedures

The study was approved by NASA Institutional Review Board (HRI-325) and took place in the Fatigue Countermeasures Laboratory at NASA Ames Research Center. Participants signed a consent form prior to participating in the study and were financially compensated for their participation. All participants reported to be right-handed. Participants had no sleep disorders as assessed by questionnaires and none of them traveled across time zones within three months of participating in the study. They were asked to complete a series of background questionnaires describing their sleep/fatigue history and habits. Before coming to the study they were required to spend 8.5 h in bed each night for two weeks, with a regular bed and wake time. Compliance with the study protocol was verified using actigraphy and daily time-stamped sleep logs. Following the at-home, sleep satiation portion of the study, we asked participants to come to the laboratory for the constant routine sleep deprivation study.

On the day of the experiment, participants arrived in the laboratory approximately two hours after their habitual wakeup time. Throughout the entire protocol, participants maintained a constant posture in bed, under < 15 lx of light, and were kept awake for at least 24 h. Ambient temperature was maintained at 22 °C. In order to dampen the influence of metabolism, daily calorie intake was calculated for each participant and meals were sub-divided into hourly isocaloric snacks. Participants were administered the five-minute PVT-192 and the five-minute NASA-PVT every two hours throughout the study, separated by 5-min break when participants were asked to provide a saliva sample. To eliminate order effects the presentation of the tasks was counterbalanced within subjects.

2.2. Equipment

The original PVT-192 was designed as a hand-held test that delivers a three-mm visual stimulus (a reaction-time counter), which requires the participant to make a pushbutton response within 1.5 s with an inter-stimulus interval varying from 2 to 10 s (Dinges and Powell, 1985).

The NASA-PVT was developed for a touchscreen PDA and was implemented on 5th generation, 32-GB Apple iPod (Apple Inc., Cupertino, CA), and running IOS 9.3.5. The development of the NASA-PVT followed all the features of the original PVT-192 described above. The

NASA-PVT was a five-minute visual PVT but it can be adjusted to any desired length. As with the traditional PVT, the inter-stimulus interval varied from 2 to 10 s randomly distributed following a rectangular distribution. The display of the stimulus was designed to look similar to the PVT-192 (red stimulus represented by a milliseconds counter in a small rectangular box on a black background). Left and right areas were predefined on the screen to serve as left or right buttons and handedness can be selected at the beginning of the test according to individual preference. The subjects received immediate feedback with the RT displayed for a second after each response was made. As with the PVT-192, if participants pressed the button too soon a "false start" appeared on the display. If participants pressed the non-dominant side of the screen an "error" message appeared on the screen. The NASA-PVT is administered with the device held in landscape orientation and participants using the thumb from their dominant hand to respond to the stimuli. This was based on results from a previous study that showed that reaction times were faster and lapses were fewer when the device was held in the landscape orientation compared to portrait orientation, and with the landscape orientation also being preferred by the participants (Arsintescu et al., 2017). A single device can be used to administer the NASA-PVT to multiple subjects during an experiment (identifiable by study code and id). The iPod was set in airplane mode and the clock time was not visible during testing. To avoid the effects of light levels produced by iPod on nocturnal melatonin levels, the brightness of the iPod was set at minimum throughout the entire experimental period (Wood et al., 2013).

2.3. Outcome metrics

The following metrics were assessed in our analyses: (1) mean 1/RT (reciprocal response time or response speed), (2) number of lapses – the cumulative number of reaction times exceeding 500 ms. The number of lapses is a valid indicator of the level of fatigue existing at the time of the test and represents lapses of attention (Dinges and Kribbs, 1991). (3) Optimum response times – the fastest 10% of response times for all trials (fastest 10% RT),which indicates the best performance a participant is capable of producing and (4) cognitive slowing – the slowest 10% of reciprocal response times for all trials (slowest 10% 1/RT), which indicates the vigilance response slowing. For mean 1/RT and slowest 10% 1/RT, a reciprocal transformation was applied to the raw data in accordance with standard methodology (Dinges and Kribbs, 1991). This procedure significantly decreases the influence of long lapses and emphasizes slowing in the optimum and intermediate ranges of responses (Dinges et al., 1987). A PVT response was considered valid if RT was > 100 ms. Consistent with previous studies, responses with an RT < 100 ms were counted as false starts (Basner and Dinges, 2011).

2.4. Latency of the device

In order to determine the latency of the device we recorded a separate session of 5-min PVT using a high-speed video camera recording 500 frames/second. The apparatus and the procedures for measuring device latency are described in Arsintescu et al. (2017). We found the mean device latency for this application was 77.42 ms (SD = 16.77). The device latency was subtracted from each PVT trial before PVT analyses in order to maximize the accuracy of PVT results.

A series of mixed effects ANOVA with two within subjects factors (Time awake x PVT type) was conducted for each PVT metric. Additional trend analyses were conducted to examine how the two types of PVT change with time awake. The Shapiro-Wilk test was used to test the normality of the model residuals for each variable (mean 1/RT; mean fastest 10% RT, mean slowest 10% 1/RT, and lapses). We found that the model residuals of all variables except lapses were normally distributed. As a result we transformed the lapses using the ln (x + 1) formula and found that its residuals were normally distributed. Therefore, for our analyses we used the transformed lapses. IBM SPSS

Statistics 23 was used for all analyses.

3. Results

Ten participants (5 males, 5 females) participated in the study. The participants were between 19 and 38 years of age (M = 25.1, SD = 6.17) and in good health. The mean time in bed for the at-home, sleep satiation portion of the study was M = 8.48 (SD = 0.25) and the mean sleep time was M = 7.04 (SD = 0.44). The average start time of the testing was 12:28 (SD = 1:44) and the average finish time was 6:57 (SD = 0:59).

For the mean 1/RT there was a main effect of time awake ($F(10, 190) = 16.85$, $p < 0.001$) for both the PVT-192 and the NASA-PVT. Trend analyses showed a significant linear increase over time for the PVT-192 ($F(10, 90) = 7.60$, $p < 0.001$) and for the NASA-PVT ($F(10,90) = 8.94$, $p < 0.001$). There was also a main effect of PVT ($F(1, 190) = 11.68$, $p = 0.001$).

Analysis of lapses indicated a main effect of time awake ($F(10, 190) = 19.12$, $p < 0.001$) and a main effect of PVT ($F(1190) = 6.56$, $p < 0.05$). Both PVTs showed a significant linear increase of lapses over time ($F(10, 90) = 11.30$, $p < 0.001$ for the PVT-192 and $F(10,90) = 7.57$, $p < 0.001$ for the NASA-PVT).

For the mean fastest 10% RT, the analyses indicated a main effect of time awake ($F(10, 190) = 7.71$, $p < 0.001$) a main effect of PVT ($F(1, 190) = 4.17$, $p < 0.05$). Both PVTs showed a significant linear increase over time ($F(10, 90) = 3.57$, $p = 0.001$ for PVT-192 and $F(10,90) = 4.37$, $p < .001$ for NASA-PVT).

For the slowest 10% 1/RT, the analyses indicated a main effect of time ($F(10, 190) = 21.59$, $p < 0.001$) and a main effect of task ($F(1, 190) = 8.00$, $p < 0.01$). Both PVTs showed significant linear increase over time ($F(10, 90) = 12.45$, $p < 0.001$ for the PVT-192 and $F(10,90) = 8.81$, $p < 0.001$ for the NASA-PVT) A visual description of the findings is provided in Figs. 1 and 2.

4. Discussion

We developed a PVT application containing the same features as the original PVT-192 to use on touchscreen devices. In order to validate the
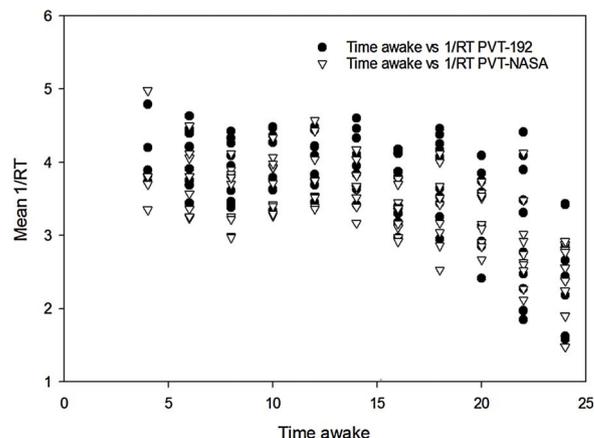
Fig. 2. Mean reciprocal response time (1/RT) for the PVT-192 and the NASA-PVT by time awake for each individual participant. Circles indicate PVT-192 responses for each individual over time and triangles indicate NASA-PVT responses.

PVT application against the original PVT-192 we conducted a laboratory experiment in which the participants were awake for approximately 24 h and asked to take both the NASA-PVT and the PVT-192 every two hours.

We found that the NASA-PVT follows the same shape and pattern and was similar to the PVT-192 on all outcome metrics after the latency cutoff was applied. The reaction times were slower and the lapses were higher as time advanced on both the NASA-PVT and the PVT-192. Overall, there was a decline in performance after 16 h awake, which coincided with the time the participants were usually going to bed and the worst performance occurred after 24 h of wakefulness (between 6:00–8:00 am) for both PVTs. Although the metrics of NASA-PVT showed the same pattern as those of PVT-192 there were significant differences between the two PVTs on mean 1/RT, slowest 10% 1/RT and fastest 10% RT. The fastest 10% RT, response speed, and slowest 10%1/RT were each slightly faster for the PVT-192. This could be due to operational differences of the two devices. For example, in previous studies we found that the latency of touchscreen devices was variable
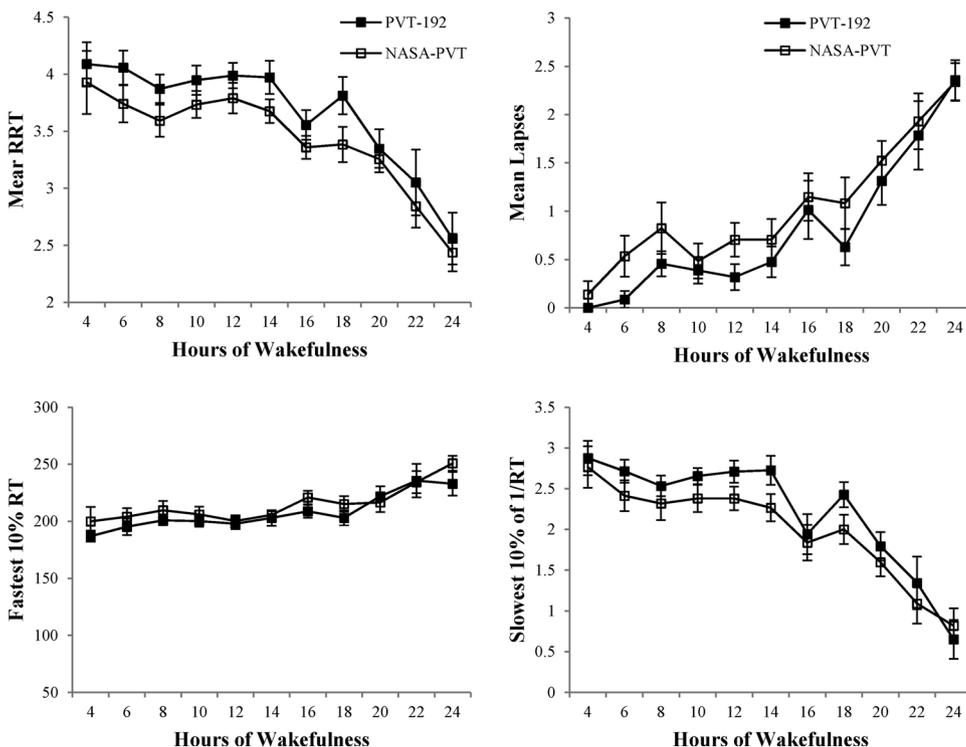
Fig. 1. Mean (± SEM) reciprocal response time (1/RT), lapses, fastest 10%RT, slowest 10% 1/RT for the PVT-192 and NASA-PVT by hours of wakefulness during the 24 h of sustained wakefulness.

among individuals and between tests (Arsintescu et al., 2017). Although the sources of variability given by the device cannot be controlled, it would be prudent for researchers using touchscreen devices to minimize all other potential sources of variability, such as through turning off other applications and Wi-Fi, and using the same device, operating system, and PVT application for all tests in a given study.

There are a multitude of reaction time tasks in the app store for which there is no information on their validation or their development. Therefore, researchers should exercise caution when using such tests on various mobile devices. Not every PVT application has been designed and validated to be used for detecting fatigue due to sleep loss and they cannot be used on any touchscreen device without knowing the latency of the device.

The NASA-PVT was designed to retain all the characteristics of the PVT-192. There are other PVTs available for mobile devices, which have been validated in laboratory studies, however, each of these PVTs have different characteristics. For example, Honn et al. (2015) evaluated a 5-min PVT on a 32 GB iPod which had some characteristics consistent with the NASA-PVT and PVT-192 (i.e., ISI between 2 and 10 s, immediate feedback at the end of each trial, presence of FS). However, that PVT differed from the NASA-PVT in that the PVT evaluated by Honn displayed a black bulls-eye on a white background as a stimulus. It also had an abort button present on the screen during the test and provided feedback on subjects' performance at the end of the test. Similarly, Kay et al. (2013) developed a 5-min PVT on an android phone with similar characteristics to the NASA-PVT, but they used a high contrast checkboard pattern on a white background as a stimulus. Other versions of PVTs have been developed for android phones or iPad mini using a three-minute task duration (Grant et al., 2017), which requires a different ISI and yields a different effect size relative to the five and ten minute versions of the task (Basner et al., 2011). The original PVT-192 has been validated in hundreds of studies investigating fatigue due to sleep loss. We aimed to make the NASA-PVT as similar to the original PVT-192 as possible in an effort to minimize the variables that might alter an individual's response profile relative to the PVT-192. These differences may be important depending on how a user intends to implement the PVT. For example, it may be desirable to use the NASA-PVT for research studies due to the similarity of our task to the PVT-192, however, it may be desirable to use a PVT that provides feedback to the user in situations where individuals are using the PVT to help understand and manage their fatigue level.

The NASA-PVT was designed to be used on Apple devices, allowing it to be installed on small testing devices such as the iPod that can be easily carried around in a field setting. Currently the application requires knowing the device latency in advance and the results analyzed after the device latency is removed from the raw data. Although the device latency can be embedded in the software design and be removed automatically when collecting data, we are reluctant to do so until the issues of device variability becomes clearer and accounted for accurately.

Our results are limited to acute sleep deprivation and therefore it is unclear how the two devices vary during chronic sleep restriction and circadian misalignment, although we have no reason to expect that they will show different results. This study was also limited by its small sample size, which may limit the generalizability of the results. Another potential limitation was that we compared the NASA-PVT against the 5-min PVT-192 instead of the original 10-min PVT. The 10-min PVT has been shown to be more sensitive in showing increased number of lapses with increased wakefulness and higher reaction times when was compared to 5-min PVT (Loh et al., 2004; Roach et al., 2006). In our study we found faster reaction times and no difference in lapses for the 5-min PVT-192 compared to NASA-PVT. Therefore, it is possible that in a

comparison with the 10-min PVT, the differences in mean outcomes between the two PVTs would be reduced.

In conclusion, our findings suggest that the NASA-PVT developed for touchscreen devices was sensitive to the effects of sleep loss and could be a reliable tool to measure performance decrements due to fatigue. In particular, the application could easily be used in field studies (e.g., aviation, astronauts during space missions) due to its easy implementation on touchscreen devices.

Funding

References

Arsintescu, L., Mulligan, J.B., Flynn-Evans, E.E., 2017. Evaluation of a psychomotor vigilance task for touch screen devices. Hum. Factors 59 (4), 661–670.

Basner, M., Dinges, D.F., 2011. Maximizing sensitivity of the Psychomotor Vigilance Test (PVT) to sleep loss. Sleep 34, 581–591.

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 Astronaut. 69 (11-12), 949–959.

Dinges, D.F., Kribbs, N.B., 1991. Performing while sleepy: effects of experimentally induced sleepiness. In: Folkard, S., Monk, T.H. (Eds.), Sleep, Sleepiness and Performance. John Wiley & Sons, Chichester, London, pp. 97–128.

Dinges, D.I., Powell, J.W., 1985. Microcomputer analysis of performance on a portable, simple visual RT task sustained operations. Behav. Res. Methods Instrum. Comput. 17, 652–655.

Dinges, D.F., Orne, M.T., Whitehouse, W.G., Orne, E.C., 1987. Temporal placement of a nap for alertness: contributions of circadian phase and prior wakefulness. Sleep 10, 313–329.

Dorrian, J., Lamond, N., van den Heuvel, C., Pincombe, J., Rogers, A.E., Dawson, D., 2006. A pilot study of the safety implications of Australian nurses' sleep and work-hours. Chronobiol. Int. 23, 1149–1163.

Grant, D.A., Honn, K.A., Layton, M.E., Riedy, S.M., Van Dongen, H.P.A., 2017. 3-minute smartphone-based and tablet-based psychomotor vigilance tests for the assessment of reduced alertness due to sleep deprivation. Behav. Res. Methods 49, 1020–1029.

Honn, K.A., Riedy, S.M., Grant, D.A., 2015. Validation of a portable, touch-screen psychomotor vigilance test. Aerosp. Med. Hum. Perform. 86 (5), 428–434.

Kamimori, G.H., Johnson, D., Thorne, D., Belenky, G., 2005. Multiple caffeine doses maintain vigilance during early morning operations. Aviat. Space Environ. Med. 76, 1046–1050.

Kay, M., Rector, K., Consolvo, S., Greenstein, B., Wobbrock, J.O., Watson, N.F., Kientz, J.A., 2013. PVT-touch: adapting a reaction time test for touchscreen devices. Proceedings of the 2013. In: 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops. PervasiveHealth. pp. 248–251.

Lamond, N., Dawson, D., Roach, G.D., 2005. Fatigue assessment in the field: validation of a hand-held electronic psychomotor vigilance task. Aviat. Space Environ. Med. 76, 486–489.

Lamond, N., Jay, S.M., Dorrian, J., Ferguson, S.A., Roach, G.D., Dawson, D., 2008. The sensitivity of a palm-based psychomotor vigilance task to severe sleep loss. Behav. Res. Methods 40, 347–352.

Lee, I.-S., Bardwell, W.A., Ancoli-Israel, S., Dimsdale, J.E., 2010. Number of lapses during the psychomotor vigilance task as an objective measure of fatigue. J. Clin. Sleep Med.: JCSM 6, 163–168.

Loh, S., Lamond, N., Dorrian, J., Roach, G., Dawson, D., 2004. The validity of psychomotor vigilance tasks of less than 10-minute duration. Behav. Res. Methods Instrum. Comput. 36 (2), 339–346.

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

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. Behav. Res. Methods 37, 111–118.

Van Dongen, H.P., Maislin, G., Mullington, J.M., Dinges, D.F., 2003. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. Sleep 26, 117–126.

Veksler, B.Z., Gunzelmann, G., 2017. Functional equivalence of sleep loss and time on task effects in sustained attention. Cognit. Sci.

Wood, B., Rea, M.S., Plitnick, B., Figueiro, M.G., 2013. Light level and duration of exposure determine the impact of self-luminous tablets on melatonin suppression. Appl. Ergon. 44 (2), 237–240.