2021-01

Wrist-Worn Psychomotor Vigilance Task Device Validation Study

Shattuck, Nita Lewis; Matsangas, Panagiotis

Monterey, California. Naval Postgraduate School

http://hdl.handle.net/10945/69155

NPS-OR-21-001

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

WRIST-WORN PSYCHOMOTOR VIGILANCE TASK DEVICE

VALIDATION STUDY

by

Nita Lewis Shattuck, Ph.D. and Panagiotis Matsangas, Ph.D.

January 2021

Distribution Statement A

Approved for public release; distribution is unlimited

Prepared for: Naval Advanced Medical Development Program; Naval Medical Research Center; 503 Robert Grant Avenue, Silver Spring, MD 20910.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE ***(DD-MM-YYYY)***	**2. REPORT TYPE**	**3. DATES COVERED** ***(From-To)***
01-15-2021	Technical Report	October 2015 – September 2019

4. TITLE: Wrist-Worn Psychomotor Vigilance Task Device Validation Study

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S): Nita Lewis Shattuck and Panagiotis Matsangas

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES):
Operations Research Department, Naval Postgraduate School; Monterey, CA 93943

8. PERFORMING ORGANIZATION REPORT NUMBER
NPS-OR-21-001

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES):
Naval Advanced Medical Development Program; Naval Medical Research Center; 503 Robert Grant Avenue, Silver Spring, MD 20910.

10. SPONSOR/MONITOR'S ACRONYM(S)
NAMD

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES
The views expressed in this report are those of the author(s) and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

14. ABSTRACT
Collecting reliable human performance data in military operational environments is an ongoing challenge. A major issue of concern is identifying systems that can capture human performance metrics in a reliable and valid manner in the field. One task that has been used extensively in the laboratory - and more recently in field settings - is the Psychomotor Vigilance Test (PVT). Since 2008 the PVT has been included as a feature on a wrist-worn actigraph (AMI, Inc.), thereby enabling researchers to easily administer the PVT in the field. The current project has three aims, a) to validate the 3-minute PVT which is currently embedded in the AMI actigraph, b) to explore the utility of other devices for field use, and c) to provide recommendations for the collection of PVT data in the military operational environments. In our studies, we found that when the screen backlight is illuminated, the results of the 3-minute PVT on the AMI actigraph are comparable to those from the laptop PVT. These findings demonstrate that the 3-minute PVT on the AMI actigraph is a valid alternative to the 3-minute laptop-based PVT for field assessment. Given the widespread use of hand-held devices with a touch screen interface, we also tested a PVT application on a representative touch screen device. Our findings were disappointing, showing that the hand-held touch screen PVT system was not comparable to the validated PVT. In particular, the touch screen PVT introduced a large constant bias as well as a proportional bias that decreased the range of response speed. These findings raised the question of what the appropriate user interface should be for a field-grade PVT system. Herein, we present a method we have developed and refined over multiple years to prepare PVT data collected in field setting for analysis. This method reduces bias due to missing data and artifacts from external disturbances by assessing the quality of the PVT data on three levels: the raw response level, the aggregated (trial) level, and the participant level.
In brief, our experience of collecting PVT data, combined with findings from this current 3 year effort, suggests that when administering the PVT in field settings, it is best if the PVT is embedded in a wearable device such as the AMI device. While other types of devices (e.g., smartphones, iPods, tablets, etc.) can be used for PVT administration, they have important constraints which should be taken into consideration.

15. SUBJECT TERMS
Psychomotor vigilance task (PVT), field studies

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			
Unclassified	Unclassified	Unclassified	Unclassified	80	Nita Lewis Shattuck
					19b. TELEPHONE NUMBER (831) 656-2281

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

THIS PAGE INTENTIONALLY LEFT BLANK

NAVAL POSTGRADUATE SCHOOL
Monterey, California 93943-5000

The report entitled "*Wrist-Worn Psychomotor Vigilance Task Device Validation Study*" was funded by, and prepared for, Naval Advanced Medical Development Program; Naval Medical Research Center; 503 Robert Grant Avenue, Silver Spring, MD 20910.

Further distribution of all or part of this report is authorized.

This report was prepared by:

Nita Lewis Shattuck, Ph.D.
Professor

Panagiotis Matsangas, Ph.D.
Lecturer

Reviewed by:

Matt Carlyle
Chairman
Department of Operations Research

Released by:

Jeffrey D. Paduan
Dean of Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Collecting reliable human performance data in military operational environments is an ongoing challenge. A major issue of concern is identifying systems that can capture human performance metrics in a reliable and valid manner in the field. One task that has been used extensively in the laboratory - and more recently in field settings - is the Psychomotor Vigilance Test (PVT). Since 2008 the PVT has been included as a feature on a wrist-worn actigraph (AMI, Inc.), thereby enabling researchers to easily administer the PVT in the field. The current project has three aims, a) to validate the 3-minute PVT which is currently embedded in the AMI actigraph, b) to explore the utility of other devices for field use, and c) to provide recommendations for the collection of PVT data in the military operational environments.

In our studies, we found that when the screen backlight is illuminated, the results of the 3-minute PVT on the AMI actigraph are comparable to those from the laptop PVT. These findings demonstrate that the 3-minute PVT on the AMI actigraph is a valid alternative to the 3-minute laptop-based PVT for field assessment.

Given the widespread use of hand-held devices with a touch screen interface, we also tested a PVT application on a representative touch screen device. Our findings were disappointing, showing that the hand-held touch screen PVT system was not comparable to the validated PVT. In particular, the touch screen PVT introduced a large constant bias as well as a proportional bias that decreased the range of response speed. These findings raised the question of what the appropriate user interface should be for a field-grade PVT system.

Herein, we present a method we have developed and refined over multiple years to prepare PVT data collected in field setting for analysis. This method reduces bias due to missing data and artifacts from external disturbances by assessing the quality of the PVT data on three levels: the raw response level, the aggregated (trial) level, and the participant level.

In brief, our experience of collecting PVT data, combined with findings from this current 3 year effort, suggests that when administering the PVT in field settings, it is best if the PVT is embedded in a wearable device such as the AMI device. While other types of

devices (e.g., smartphones, iPods, tablets, etc.) can be used for PVT administration, they have important constraints which should be taken into consideration.

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGEMENTS

A number of individuals were involved in this project:

- NPS faculty: Frances Greene, Ph.D.
- National Research Council post-doctoral associate: Dr. Sjörs Dahlman
- NPS students: LTJG Selina Benavides, MAJ Rachel Gonzales, LCDR Imran Sikandar
- Summer interns: Raymond Garcia, Cynthia Jones, Jenny Lafayette, Alexander Laney, Arjun Mayur, Katherine Mortimore, Gabriella Nagy, Christopher Paghasian, Alexis Sakoda, Sierra Smith, Joshua Strubel, Amanda Vanecko, Sophia Winland, Candace Wong, Isaac Yoon

Also, the authors thank Heather Clifton, Ph.D. for reviewing and proofreading the manuscript.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. BACKGROUND

Many tasks in military operations include a vigilance component that requires the operator to pay attention for prolonged periods of time to detect the infrequent occurrence of critical events. It is well-documented that human ability to sustain focused attention deteriorates over time. This phenomenon is further exacerbated by sleepiness and fatigue, both common problems in military environments.

One task used widely in laboratory and field studies to assess sustained attention is the visual Psychomotor Vigilance Task (PVT) (Dinges & Powell, 1985). The PVT is a simple reaction time task in which participants press a response button when a stimulus appears on a display screen. PVT performance is not only affected by sleep loss but is also sensitive to circadian rhythmicity (Dinges et al., 1997; Doran, Van Dongen, & Dinges, 2001; Durmer & Dinges, 2005; Jewett, Dijk, Kronauer, & Dinges, 1999; Wyatt et al., 1997). It has the additional advantage of being relatively impervious to practice effects (Basner et al., 2018). Taken together, these characteristics make the PVT an appealing candidate for field-based military performance research and potential fitness-for-duty assessments. In fact, the PVT is considered the gold standard for assessing the neurobehavioral effects of sleep loss and circadian misalignment, both commonly found in operational settings.

The original version of the PVT is 10 minutes in duration and is administered using a dedicated, validated device (Dinges & Powell, 1985) or a laptop. This duration is problematic in operational field studies for several reasons. First, many individuals are already heavily tasked and pulling them away from their actual work to engage in an artificial 10-minute task and adding to their already heavy workload is not well received. Second, providing a laptop to each individual participant in an operational study is not cost-effective but centrally locating the testing device in a library or common access area requires participants to go out of their way to participate—again, adding to the requirements of an already overbooked schedule. Lastly, many operational environments will not allow the introduction of laptops or other devices due to security concerns. Considering these issues, it is no surprise that the original 10-minute PVT is considered excessively cumbersome by most operators, and logistically difficult or inappropriate to incorporate into field-based studies by many researchers.

One alternative to the laptop-based 10-minute PVT is the 3-minute version embedded in a wrist-worn actigraphy device (Ancoli-Israel et al., 2003; Mullaney, Kripke, & Messin, 1980). While wearing the actigraph to assess activity and sleep patterns, participants can also perform the PVT without leaving the workplace. In recent years, researchers at the Naval Postgraduate School have used the integrated wrist-worn device (WWD) with a great deal of success in multiple shipboard studies (for example, Shattuck, Matsangas, & Brown, 2015; Shattuck, Matsangas, & Powley, 2015; Shattuck, Matsangas, & Waggoner, 2014; Shattuck, Waggoner, Young, Smith, & Matsangas, 2014). Participant feedback has shown high acceptance of the WWD PVT in terms of utility and ease of use. Additionally, participant compliance is higher and the attrition rate is reduced when using the WWD version compared to the laptop PVT.

Another alternative is to administer the PVT using a hand-held device (HHD). Given the widespread use of HHDs, their relative affordability, and their portability, several research efforts have explored the use of an HHD PVT (Arsintescu et al., 2019; Arsintescu, Mulligan, & Flynn-Evans, 2018; Grant, Honn, Layton, Riedy, & Van Dongen, 2017; Honn, Riedy, & Grant, 2015; Kay, Grandner, et al., 2013; Kay, Rector, et al., 2013; Ocano, Watson, Kay, Kientz, & Grandner, 2017). These studies have utilized a number of different devices (e.g., Apple iPad, Apple iPod, Samsung Galaxy) and multiple versions of the PVT (3-minute, 5-minute, 10-minute). While some differences between the HHD and laptop PVTs in terms of reaction times, false starts, and attention lapses, overall, the HHD PVT has been found to be sensitive to the effects of fatigue, thereby making it a potentially useful tool in operational settings.

An important PVT parameter is the inter-stimulus interval (ISI), which denotes the period between the last response and the appearance of the next stimulus. From a statistical perspective, a number of data points are needed to reliably assess the alertness state of an individual. However, vigilance is a phenomenon associated with infrequently occurring events. Therefore, the rate of stimuli in the task must be kept low in order to better represent real-world vigilance scenarios. When we consider the frequency of events that occur in the operational environment, the ISI would be set unrealistically low for the original 10-minute duration of the PVT. This problem is further magnified when the task duration is shortened to 3 minutes. Therefore, the duration of the ISI is not trivial. The ISI typically used in the 10-minute version of the PVT is 2 to 10 seconds. When Basner and colleagues introduced the 3-minute PVT, they increased the signal rate by reducing the ISI 1 to 4 seconds (Basner, Mollicone, & Dinges, 2011; Basner & Rubinstein, 2011). The authors noted that this increase in signal rate partially compensated for the reduction in the number of responses due to decreasing the trial duration from 10 to 3 minutes. However, this change in ISI, in conjunction with the trial duration, led to faster responses, increased false start rates, and decreased lapse frequency in the brief PVT compared to the original 10-minute PVT (Basner et al., 2011).

B. PROBLEM STATEMENT

In the military operational environment, it is important to understand how human performance varies as a function of issues such as stress, fatigue, manning and watchstanding. Collecting reliable in the operational environment, however, continues to be a challenge. One issue of concern is finding field-worthy systems that can capture human performance metrics in a reliable and valid manner.

The second issue is the operational (external) validity of the test to be used to capture human performance. Given that this project is focused on collecting PVT data in operational conditions, we decided to assess the validity of the short, 3-minute, PVT using both the larger and smaller ISIs. It is the authors' opinion, however, that a larger ISI denoting a less frequent task is more representative of the operational environment.

C. STUDY AIMS AND OBJECTIVES

The current effort has three aims. The first aim is to validate the 3-minute PVT which is currently embedded in the AMI actigraphs. The second aim is to explore the utility of other devices for field use. Lastly, the third aim is to provide recommendations for the collection of PVT data in military operational environments.

Overall, this project investigated the feasibility of DoD-wide individual use of the PVT in the AMI wrist-worn device for fitness-for-duty applications. The multiyear project was conducted in two phases.

1. Phase 1 (2014 – 2015)

In fiscal year 2015, our first study to validate the wrist-worn PVT. This study allowed us to determine if there were differences in PVT performance due to the test devices and/or ambient conditions in order to isolate those as potential sources of human performance variance. Specifically, the study focused on the following objectives:

- Compare the 3-minute WWD with ISI between 2 to 10 seconds with the 3-minute laptop PVT with ISI between is 1 to 4 seconds.
- Assess the effect of the backlight feature provided on the WWD.
- Assess the effect of ambient lighting on PVT performance metrics for both device types.
- Determine the magnitude of the learning effect for the 3-minute WWD PVT.

2. Phase 2 (2016 – 2018)

In Phase 2 of this effort, we focused on further assessing the utility of the 3-minute PVT with the following objectives:

- FY2016
  - Compare the 3-minute WWD PVT results with the 3-minute laptop PVT when the ISI for both devices is 1 to 4 seconds.
  - Assess the effect of the backlight feature provided in the wrist-worn device when ISI is 1 to 4 seconds.
  - Determine the magnitude of the learning effect for the 3-minute WWD PVT when ISI is 1 to 4 seconds.
- FY2017

    - Explore the availability of new WWDs which can be used for activity/sleep assessment and include an embedded version of the PVT.
    - Explore the utility of new devices which can be used for activity/sleep assessment and include an embedded version of the PVT.
    - Compare the new PVT results with the laptop PVT.
- FY2018
    - Explore the utility of new devices which can be used for activity/sleep assessment and include an embedded version of the PVT.
    - Compare the new PVT results with the laptop PVT.
    - Determine the magnitude of the learning effect for the new PVT.

D. REPORT STRUCTURE

The Methods chapter in this report describes the various tools and technologies as well as the procedures used in the experiments we performed. The Results chapter describes the findings from these experiments and includes lessons learned from the collection of PVT in the field. The Conclusions chapter provides an overview of our findings and discusses the importance of these findings in terms of using the PVT embedded in wrist-worn and other devices for fitness-for-duty applications in the DoD. The Recommendations chapter proposes potential routes for future research. This report includes the following Appendices.

- Appendix A: List of the products developed under this project.
- Appendix B: Presentation with the results of FY2015.
- Appendix C: Presentation with the interim results of FY2016.
- Appendix D: Abstract accepted to the Military Health System Symposium (MHSRS) 2017.

THIS PAGE INTENTIONALLY LEFT BLANK

II. METHODS

A. PARTICIPANTS

A total of 135 volunteers (18 to 63 years of age, 93 males) from the Naval Postgraduate School (NPS) volunteered to participate in the three experiments. Participants were screened for corrected vision, recent injuries or pain in the arms, wrists or fingers, or a diagnosis of color vision deficiency or carpal tunnel syndrome. The NPS Institutional Review Board approved the protocols used (NPS.2015.0013, NPS.2016.0018). Informed consent was obtained from all volunteers after the experimental procedures had been fully explained.

B. EQUIPMENT

1. PVT systems

Psychomotor Vigilance Task (PVT) performance data were collected using three different devices. Attributes of each device are discussed below, including test duration and the inter-stimulus interval (ISI). Metrics include reaction time (RT), response speeds, false starts, and lapses in attention, and are discussed in detail later.

a. The validated laptop PVT

The PVT (implementation version 2.0.5.9 – Pulsar Informatics Inc., Philadelphia, PA) was installed on individually calibrated Latitude E5420 and E6420 laptops with 14″ displays (Dell Inc., Round Rock, TX), running the Windows 7 operating system (Microsoft Inc., Redmond, WA). The characteristics of the laptop PVT are shown in Table 1.

Table 1. Attributes of the laptop PVT

Stimulus	Visual millisecond counter in a red rectangular box
Test Duration	3 minutes
ISI	Random: 1-4 seconds
Feedback	RT in milliseconds is displayed for 1s (this period is part of the next ISI)
False starts	"FS" is displayed for 1 second (this period is part of the next ISI)
Time out without a response	After 30 seconds, "OVERRUN" is displayed for 1 second with an auditory signal to alert the subject. The response is counted as valid, i.e., as a lapse with a response time of 30 seconds.
Button fails to release	"BUTTON" is displayed after the response button has not been released for 3 seconds and a signal is continuously played back until the button is released. The new ISI starts once the button is released.
Wrong Key Pressed	"ERR" is displayed for 1 second if the wrong response key is pressed.

b. The PVT embedded to a wrist-worn device

The Motionlogger Watch (Ambulatory Monitoring, Inc., Ardsley, NY) was the wrist-worn device (WWD) used in our experiments. The Motionlogger Watch had an embedded version of the PVT. The characteristics of the WWD PVT are shown in Table 2.

Table 2. Attributes of the PVT embedded in the wrist-worn device.

Stimulus	The appearance of the word "PUSH" in black digital letters presented on a low contrast LCD screen. When chosen, the red backlight will illuminate concurrently with the appearance of the word "PUSH".
Test Duration	3 minutes
ISI	Random: 1-4 seconds or 2-10 seconds
Feedback	RT in milliseconds is displayed for 1 second
False starts	"FS" is displayed for 1 second
Time out without a response	After 60 seconds, "DONE" is displayed for 1 second and the test session is closed. The response is not counted as valid.
Button fails to release	Device does not address this contingency. As soon as the button is pressed, the response is counted. Upon the next stimulus, the button is not recognized as depressed.
Wrong Key Pressed	All available buttons are valid

A note is appropriate for the screen backlight feature mentioned in Table 2. The standard PVT stimulus is the word "PUSH" in black digital letters, which is presented on a low contrast LCD screen. The intensity of the ambient light, therefore, may affect the PVT performance on the WWD. When the screen backlight is on, however, the red backlight turns on concurrently with the "PUSH" stimulus. The light helps illuminate the word "PUSH" but also introduces a secondary visual cue, thereby enhancing the stimulus. The WWD screen when the backlight is off and on is shown in Figure 1.

Figure 1. Wrist-worn device (WWD) with the screen backlight feature off (left) and on (right).

c. The PVT-Touch embedded in a tablet

The PVT-Touch (Kay, Grandner, et al., 2013; Kay, Rector, et al., 2013) was installed on two Samsung Galaxy Note 8.0 GT-N5110 (Samsung Electronics Co., Ltd., Suwon, South Korea; Figure 2) tablets that served as the touch-screen hand-held devices (HHDs).

Figure 2. The Samsung Galaxy Note 8.0 GT-N5110.

All other applications and Wi-Fi were turned off to minimize processing speed variability. Participants were instructed to use the touch down technique on the screen to respond to the PVT stimulus (Kay, Rector, et al., 2013). In contrast to other PVT studies (Arsintescu et al., 2018; Grant et al., 2017), participants were instructed to perform the HHD PVT using their dominant hand while they were holding the device on a table in a portrait orientation. We chose these instructions because we assumed that holding the HHD steady on a fixed surface in a portrait orientation would be easier when performing the PVT in a moving environment (e.g., sailors on ships).

Table 3. Attributes of the PVT-Touch.

Stimulus	A white square appearing on the black screen
Test duration	3 minutes
ISI	Random: 1-4 seconds
Feedback	The white square is replaced by the RT in white. The RT is displayed until the next stimulus.
False starts	Anticipation time displayed in red at the time of the next stimulus.

2. The Go/No Go task

The laptop-based Go/No-Go task was developed by the Naval Aviation Medical Research Unit at Dayton (NAMRU-D). This Go/No Go task includes the reaction time component seen in the PVT but also introduces a response inhibition task. Individuals performing the Go/No-Go task are instructed to respond to a “Go” stimulus but to refrain from responding to a “No-Go” stimulus.

3. Questionnaires

In all studies, participant completed a number of questionnaires with information on demographics, sleep history for the 48 h prior to the data collection, caffeine intake prior to the data collection, and any issues that might affect participation in the experiment. Detailed information on the questionnaires used in the various data collection under this project can be found elsewhere (Matsangas & Shattuck, 2018, 2020; Matsangas, Shattuck, & Brown, 2017; Matsangas, Shattuck, Mortimore, Paghasian, & Greene, 2019).

4. Actigraphy

Sleep was assessed by wrist-worn actigraphy (the Motionlogger Watch used for the PVT) assisted by paper and pen activity logs. Data were collected in 1-minute epochs using the Zero-Crossing Mode and were scored using Action W version 2.7.2155 software. The Cole-Kripke algorithm with rescoring rules was used. Criterion for sleep and wake episodes was five minutes. The sleep latency criterion was no more than one minute awake in a 20-minute period (all values were default for this software).

C. RESEARCH DESIGN AND PROCEDURES

Three studies were conducted under this project, split into four data collection periods. All studies were experimental in nature, conducted in controlled conditions, and utilized a randomized, within-subject, repeated-measures design. For a detailed description of the methods involved in each of the experiments, refer to the corresponding publications. This section will focus on providing a general overview of the methods we used.

1. First study

In the first experiment (36 participants) we compared the 3-minute WWD PVT (ISI = 2 – 10 seconds) with the laptop PVT (ISI = 1 – 4 seconds). Figure 1 illustrates the research approach used for the first round of data collection. The study utilized a randomized, repeated-measures design with three factors. The first factor is the PVT device type (laptop - L, wrist-worn Device - WWD). The second is the red backlight (BL) feature of the wrist-worn PVT (BL=ON, BL=OFF). The third factor is ambient lighting with two levels: a low ambient lighting condition similar to twilight (2 – 3 lux), and a normal office lighting environment (300 – 400 lux). Ambient lighting was counterbalanced. Within each ambient light condition, device order was completely counterbalanced. The actual illumination levels for these two test conditions were approximately 0.3 and 55 FT-candles.

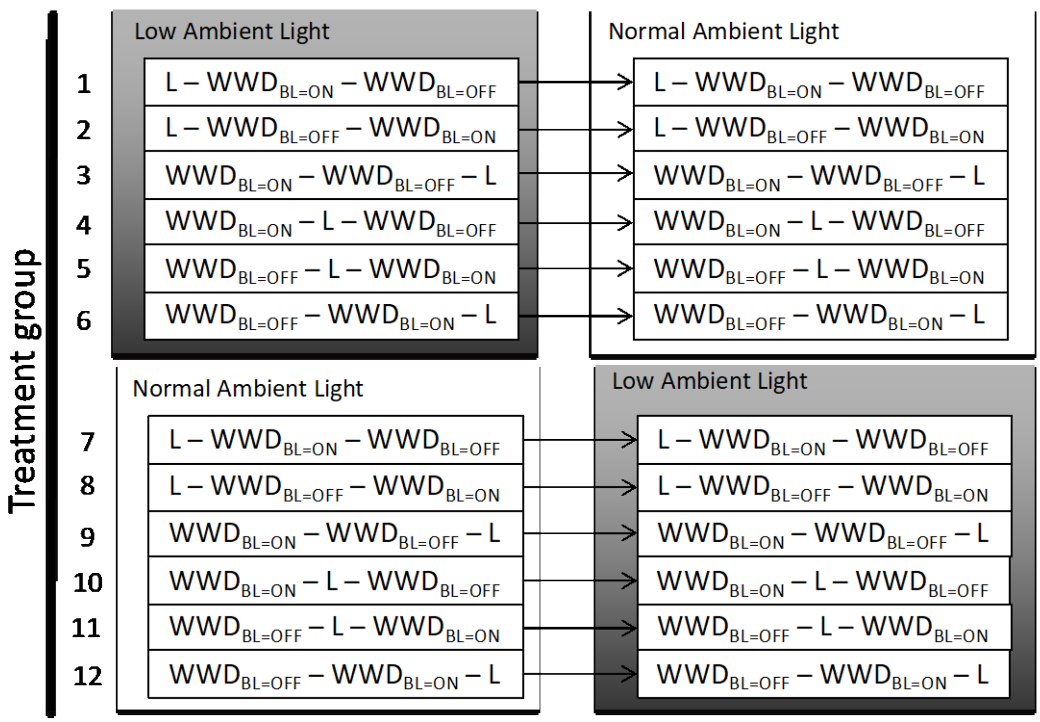


Figure 3. Study design in the first experiment (laptop – L; WWD with the backlight ON – $WWD_{BL=ON}$; WWD with the backlight OFF – $WWD_{BL=OFF}$)

Participants first completed the Study Questionnaire. They were shown how to perform the PVT and they were allowed one test trial on each device. In order to keep the reaction times as low as possible, participants were instructed to respond as soon as each stimulus appeared, but not to anticipate the stimulus to avoid a false start. Next, participants were randomly assigned to one of the twelve treatment groups of this experiment (Figure 3). All participants performed six 3-minute PVT trials. Between trials, there was a 1-minute break, whereas between ambient light conditions there was a 5-minute break. The total length of time to complete the experiment was approximately 45 minutes. While performing the tests, participants were seated and wore headphones to attenuate ambient noise. A researcher was present at all times in the experimentation room.

2. Second study

In the second experiment (58 participants split in two data collection periods) the aim was to compare the 3-minute PVT (ISI = 1 – 4 seconds) embedded in a hand-held device (HHD) with a touch-screen and the WWD PVT with the validated laptop PVT. Participants arrived at the laboratory two days before the main data collection. They were issued an actigraph and an activity log to assess their sleep patterns before and during the study (approximately three days in total). Data were collected on two consecutive days in normal office lighting environment (300 – 400 lux). After completing the Study Questionnaire, participants were shown how to perform the PVT and they were allowed one test trial with each device. Participants were instructed to respond as soon as each stimulus appeared, but not to anticipate the target because that would yield a false start.

Based on the findings of a pilot study, the experiment was split into two consecutive days to avoid participants' boredom effects and lack of focus while performing the PVT. On the first day, participants performed three 3-minute PVT trials, one on the laptop (L), one on the WWD with the screen backlight off ($WWD_{BL=OFF}$), and one on the WWD with the screen backlight on ($WWD_{BL=ON}$). Participants were randomly assigned to one of the six treatment groups, with the order of devices completely counterbalanced (Figure 4). Between trials, there was a 1-minute break. While performing the tests, participants were seated, wearing headphones to attenuate ambient noise. A researcher was present in the experimentation room, behind the participant, to monitor the study. On the second day, participants reported to the laboratory at approximately the same time as the first day of the experiment. Participants completed the ESS, and performed two 3-minute PVT trials, one on the HHD and one on the laptop in a counterbalanced order (Figure 4).

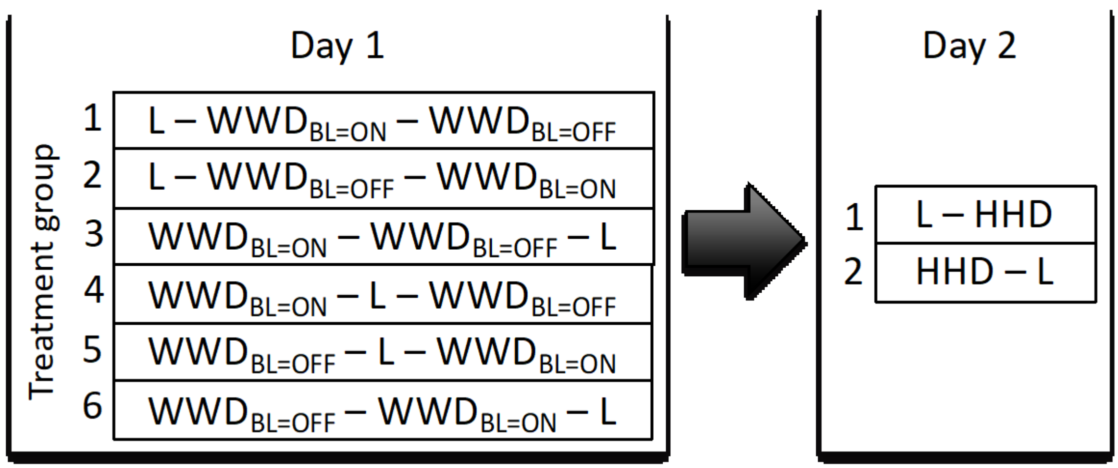


Figure 4. Study design in the second experiment (Laptop – L; WWD with the backlight ON – $WWD_{BL=ON}$; WWD with the backlight OFF – $WWD_{BL=OFF}$; Hand-held device – HHD)

3. Third study

Given the sensitivity of PVT metrics to fatigue and circadian rhythmicity, another parameter likely to be important is time of day. When using the PVT in 24-hour operational conditions to assess alertness and performance at various points in a shift or watchstanding period, the assessments will likely occur at very different phases of the circadian cycle. A systematic understanding of the effect of time of day on PVT performance is therefore necessary to validate and interpret PVT data collected during 24-hour operations.

Due to this background, the third study focused on the utility of the PVT to assess changes in performance between a morning and an afternoon data collection session. In total, 41 participants performed both the wrist-worn and the laptop 3-minute PVT during both a morning and an afternoon testing session, using both the laptop PVT and the wrist-worn PVT. The ISI for both devices was set to 1 to 4 seconds. Based on results from our previous studies, the WWD PVT was performed with the backlight feature on (i.e., the LCD screen was lit when the stimulus – the word “PUSH” – was presented).

To further validate the use of the wrist-worn PVT in an operational setting and account for the dual stimulus presented in the wrist-worn PVT by the word “PUSH” and the illumination of the backlight, we also compared the laptop and WWD versions of the PVT to the Go/No-Go task, a decision-making task similar to “friend versus foe” identification, which has been used previously in aviation research (Combs, 2018). Similar to the PVT, the Go/No-Go requires cognitive processes (sustained attention, response inhibition, and working memory) that are sensitive to the effects of total sleep deprivation and circadian phase (Chuah, Venkatraman, Dinges, & Chee, 2006; Drummond, Paulus, & Tapert, 2006; Sagaspe et al., 2012).

Participants were asked to complete the Study Questionnaire which included the Pittsburg Sleep Quality Index (PSQI) (Buysse, Reynolds, Monk, Berman, & Kupfer, 1989). They were also instructed to wear an actigraph for three days prior to the main data collection day and to complete a daily activity log. The main experiment utilized a randomized, within-subject, repeated-measures design with two factors (task, time of day). Ambient lighting was typical for an office environment (i.e., 300 – 400 lux).

Each participant arrived in the lab in the morning of the test day, received an explanation of the procedures of the experiment, and performed three training trials (one on each task: laptop PVT, WWD PVT, Go/No Go task). Next, participants performed one trial of each task with one to two minutes between trials. Task order was completely counterbalanced (i.e., there were six treatment groups in the experiment). The same procedures were followed in the afternoon session. Each participant performed tasks in the same order in the morning and in the afternoon data collection session. The morning session occurred at around 10 AM ± 1 hour, whereas the afternoon session occurred at around 3 PM ± 1 hour. Figure 5 shows the treatment groups and the order of the tasks in each group.

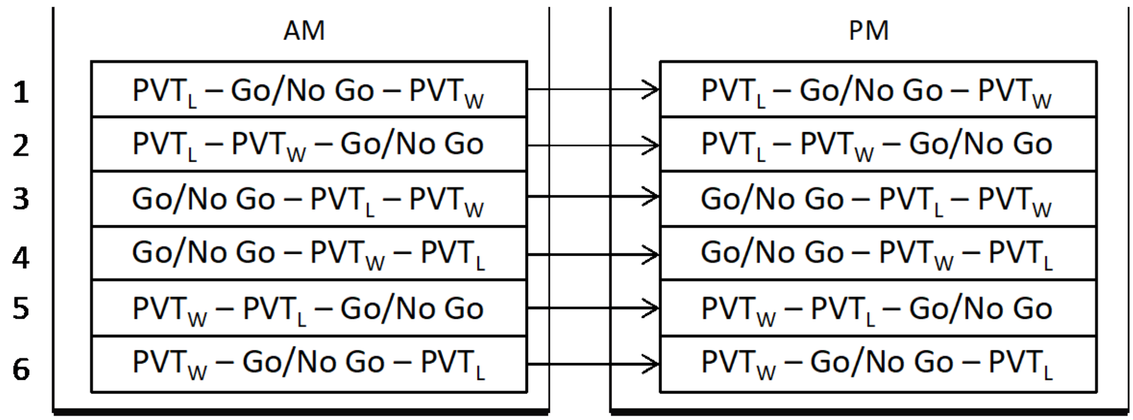


Figure 5. Study design in the third experiment (Laptop PVT– PVT_L; Wrist-worn device PVT – PVT_W)

D. STATISTICAL ANALYSIS

A PVT response was regarded as valid if the reaction time (RT) was greater than or equal to 100 milliseconds (ms) and less than 30 seconds. Responses with RTs less than 100 ms were identified as false starts (errors of commission). Lapses were defined as RTs greater than or equal to 355 ms and 500 ms. Based on the PVT metrics proposed by Basner and Dinges (2011), analysis included nine PVT metrics: mean RT, mean response speed (i.e., reciprocal reaction time, calculated as 1/RT*1000 and measured in 10^3*$msec^{-1}$), fastest 10% RT, slowest 10% 1/RT, percentage of false starts, percentage of 355 ms lapses, percentage of 500 ms lapses, percentage of 355 ms lapses combined with false starts, and percentage of 500 ms lapses combined with false starts. For all metrics, the response values were aggregated by trial.

When the Go/No-Go task was used, the task was 5 minutes in duration and 180 trials were presented. The trials consisted of 80% Go and 20% No-Go and had an ISI of 0.5 to 1 second. Responses with RTs less than 100 ms were identified as errors of commission and excluded from the RT analysis. The Go/No-Go metrics were: mean RT of correct Go trial responses, mean response speed of correct Go Trial responses (calculated as 1/RT*1000 and measured in 10^3*$msec^{-1}$), percentage of errors of omission (not responding in a Go trial), and percentage of errors of commission (false alarms – responding in a No-Go trial).

Data normality was assessed with the Shapiro-Wilk W test. Parametric and non-parametric statistical methods were used as deemed appropriate for normally and non-normally distributed data. Multiple comparisons were based on the Tukey-Kramer Honestly Significant Difference (HSD) test and the Dunn method for joint ranking, whereas pairwise comparisons were based on the t-test and the Wilcoxon Rank sums test as appropriate.

To assess the agreement between the WWD and HHD PVT systems with the validated laptop PVT, we used the Bland–Altman method (Altman & Bland, 1983; Bland & Altman, 1986, 1999). The basic Bland–Altman method was used if the mean and standard deviation of the differences between devices were the same throughout the range of measurement (Altman & Bland, 1983). The regression approach for non-uniform differences was used if the mean difference between devices was associated with the magnitude of the measurements (Bland & Altman, 1999).

An alpha level of 0.05 was used to determine statistical significance. For multiple comparisons, post-hoc statistical significance was assessed using the Benjamini–Hochberg False Discovery Rate (BH-FDR) controlling procedure (Benjamini & Hochberg, 1995; Groppe, Urbach, & Kutas, 2011) at the 0.05 level. Statistical analysis was conducted with JMP statistical software (JMP Pro; SAS Institute; Cary, NC). Normally distributed data are presented as mean ± standard deviation (M ± SD), whereas non-normally distributed data are presented as median – MD (interquartile range – IQR). Detailed information regarding the analytical approach can be found in the corresponding publications (Matsangas & Shattuck, 2018, 2020; Matsangas et al., 2017; Matsangas et al., 2019).

THIS PAGE INTENTIONALLY LEFT BLANK

III. RESULTS

A. RESULTS FROM STUDIES IN CONTROLLED CONDITIONS

In the first experiment (36 participants), we compared the 3-minute WWD PVT (ISI = 2 – 10 seconds) with the laptop PVT (ISI = 1 – 4 seconds). The cumulative distribution function (CDF) plots of the PVT reaction times (RT) showed two patterns. First, in dim light conditions, the RTs of the WWD without backlight were considerably longer than the laptop and the WWD with backlight. Second, when performing the PVT on the actigraph with the backlight on, participants tended to have faster responses compared to their performance on both the laptop and the actigraph with the backlight off. These results are shown in Figures 6 and 7.

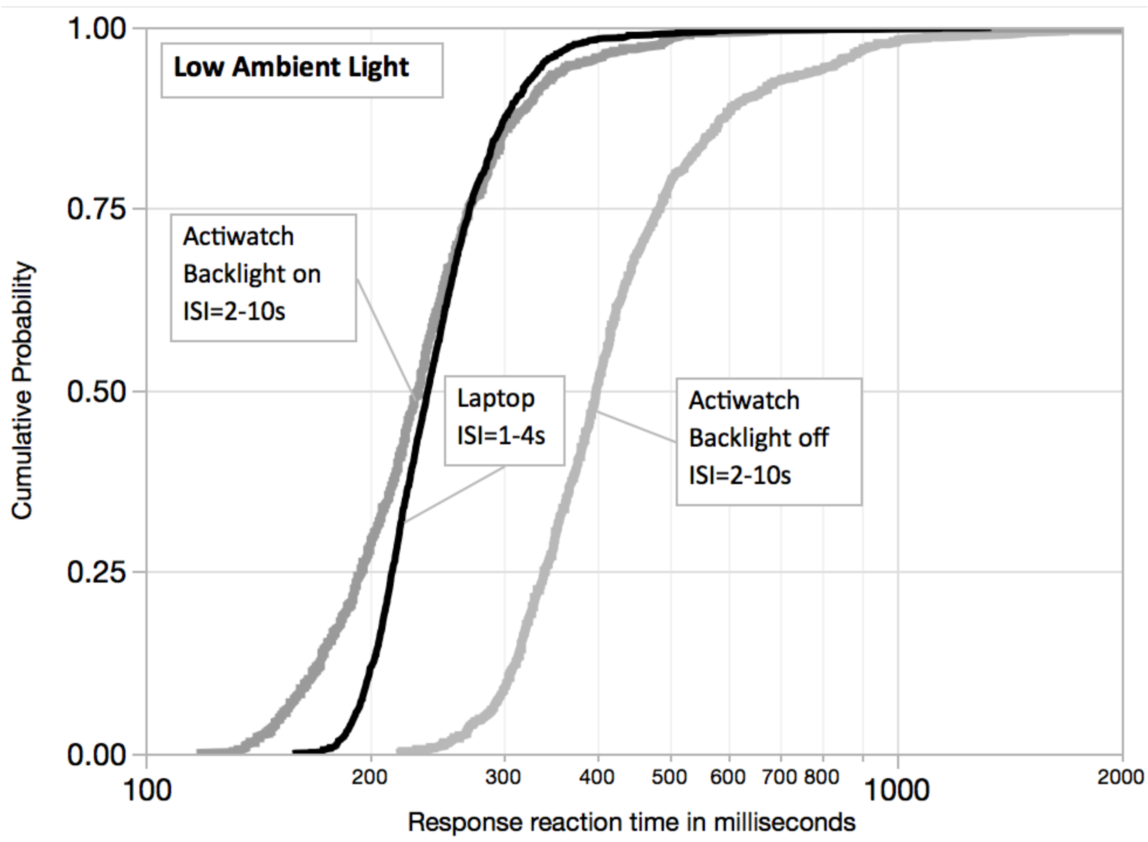


Figure 6. PVT responses in low ambient light. Wrist-worn PVT ISI=2-10s, laptop PVT ISI=1-4s.

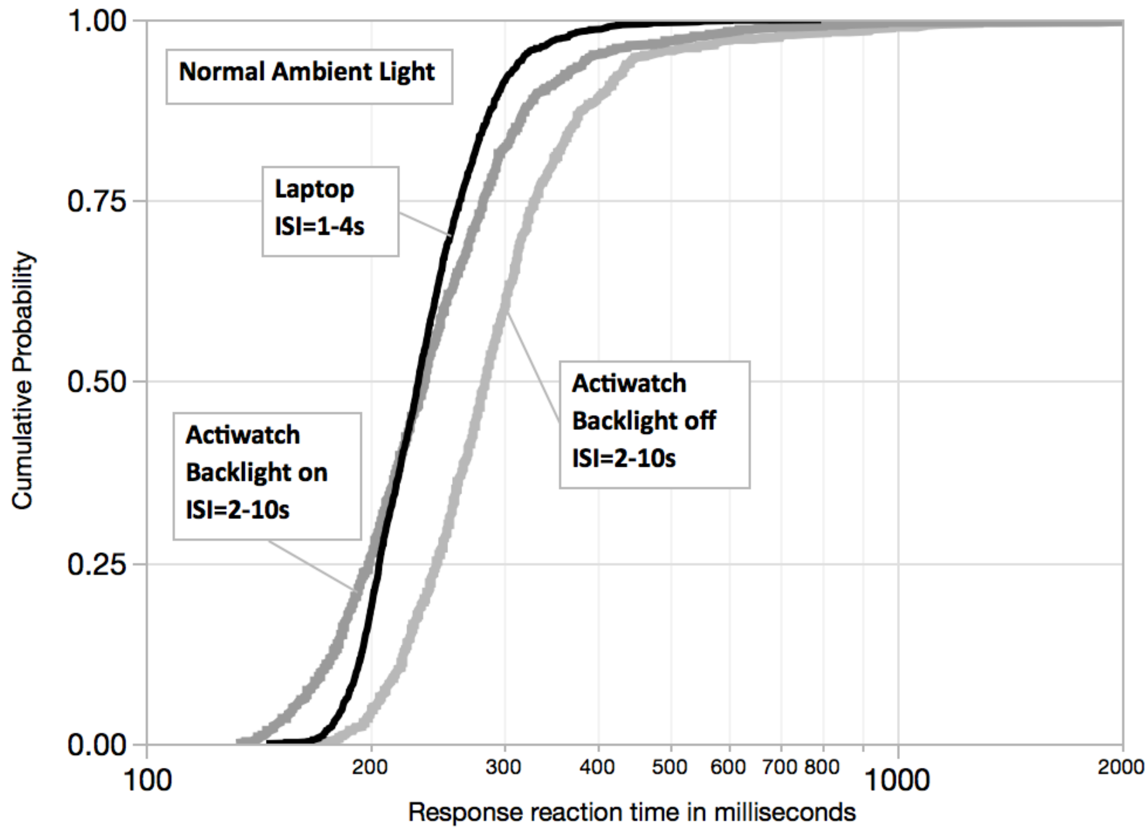


Figure 7. PVT responses in normal ambient light. Wrist-worn PVT ISI=2-10s, laptop PVT ISI=1-4s.

The remainder of our analysis will focus on PVT metrics aggregated by trial. Next, we performed a mixed effects analysis to assess the factors associated with the reciprocal reaction time (1/RT). Results showed that device type ($F(1, 176)=64.6$, $p<0.001$), ambient light condition ($F(1, 176)=30.0$, $p<0.001$), and WWD backlight feature ($F(1, 176)=292$, $p<0.001$) are all main effects associated with reciprocal reaction time. However, reciprocal RTs were not associated with the sequence of ambient light conditions ($p>0.70$).

We also assessed the median contrasts in PVT metrics between all conditions (Table 4). Columns A to C include the median contrasts within each ambient lighting condition. The median difference in RTs between the laptop and the WWD with the backlight on in low ambient light conditions (column C) was less than 10 ms which corresponds to less than 4% of the RT found in the laptop PVT. However, the absence of a backlight on the WWD has a considerable effect on the differences of PVT metrics compared to the laptop (column B). Specifically, in low ambient light conditions, the WWD shows an approximately 60% increase in mean and the fastest 10% RTs compared to the laptop. There is an approximately 76% increase in the percentage of 355 ms lapses combined with false starts, whereas the increase is approximately 11% in the percentage of 500 ms lapses combined with false starts. In normal ambient lighting condition, this pattern is less pronounced. Compared to the laptop, the WWD shows an approximately 22% increase in mean RTs, 28% increase in the fastest 10% RTs, and 8% increase in the percentage of 355 ms lapses combined with false starts. No statistically significant differences were identified in the percentage of 500 ms lapses combined with false starts.

Columns D to F include the contrasts between ambient lighting conditions. The pattern of results in column D shows that PVT performance in the WWD is sensitive to ambient lighting conditions when the backlight feature is off. In contrast, when the backlight is on (Column E), PVT performance is not affected by ambient lighting. These results are shown in Table 4. Later, we verified the patterns described so far when the WWD has ISI between 1 and 4 second. For this purpose, we used data from the second experiment to be described later.

Table 4. Median contrasts in PVT metrics by device type, WWD backlight, and ambient light conditions. Wrist-worn PVT ISI=2-10s, laptop PVT ISI=1-4s.

PVT metric	Ambient Light condition	Contrasts within each Ambient Light condition			Contrasts between Ambient Light Conditions (Low - Normal)		
		WWD with Backlight Off vs. WWD with Backlight On MD (95% CI) [A]	WWD with Backlight Off vs. Laptop MD (95% CI) [B]	WWD with Backlight On vs. Laptop MD (95% CI) [C]	WWD with Backlight Off MD (95% CI) [D]	WWD with Backlight On MD (95% CI) [E]	Laptop MD (95% CI) [F]
Mean RT, [ms]	Low	170 (152, 198)***	162 (134, 203)***	-9.77 (-21.7, -0.36)*	111 (97.5, 124)***		10.8 (6.19, 18.3)***
	Normal	54.6 (41.3, 59.8)***	64.2 (49.4, 81.0)***				
Mean 1/RT, [1000/msec]	Low	-1.99 (-2.08, -1.69)***	-1.63 (-1.83, -1.38)***	0.19 (0.02, 0.41)*	-1.03 (-1.19, -0.77)***		-0.15 (-0.22, -0.11)***
	Normal	-0.84 (-0.93, -0.66)***	-0.84 (-0.99, -0.58)***				
Fastest 10% RT, [ms]	Low	135 (120, 158)***	123 (102, 149)***	-15.9 (-27.8, -6.82)**	95.7 (81.1, 111)***		7.51 (3.66, 9.88)***
	Normal	43.2 (37.8, 54.1)***	42.6 (31.3, 53.1)***				
Slowest 10% 1/RT	Low	-1.58 (-1.72, -1.41)***	-1.40 (-1.62, -1.24)***		-0.71 (-0.94, -0.47)***		-0.16 (-0.27, -0.06)**
	Normal	-0.66 (-0.85, -0.21)***	-0.98 (-1.02, -0.69)***	-0.20 (-0.54, -0.08)*			
Lapses 355ms + False Starts, %	Low	72.6 (54.9, 81.4)***	76.3 (61.1, 86.8)***		55.3 (35.0, 70.4)***	-3.92 (-4.76, 0)**	
	Normal	5.74 (4.35, 9.05)***	8.01 (2.96, 17.0)***				
Lapses 500ms + False Starts, %	Low	13.5 (9.09, 20.0)***	10.9 (5.36, 20.0)***	-2.08 (-2.13, 0)**	9.52 (4.55, 18.0)***	0 (-4.35, 0)*	
	Normal						

Inclusion criterion: All shown contrasts are statistically significant according to post-hoc analysis with the BH- FDR controlling procedure. Empty cells denote statistically not significant results.

Statistical significance of the shown contrasts based on the Wilcoxon Signed Rank test; p-values: "*" p<0.05; "**" p<0.01; "***" p<0.001

MD (95% CI): Median (95% Confidence Interval of the median)

The analysis presented thus far was based on comparing median performance between PVT devices. Useful as it may be, though, this type of analysis does not convey all the appropriate information for comparing devices. For this reason, we further assessed differences between the WWD PVT (ISI = 2 – 10 seconds) and the validated laptop PVT (ISI = 1 – 4 seconds) using the Bland–Altman method (Altman & Bland, 1983; Bland & Altman, 1986, 1999).

The next two figures show the scatter diagrams of the RTs in the two devices. The solid black lines indicate the equality (45 degrees) lines. The WWD PVT with the backlight off produces longer reaction times (slower response speeds) compared to the laptop PVT, with the differences being most pronounced at low ambient light (Figure 8) and more modest under normal ambient lighting conditions (Figure 9). When the backlight is on, the PVT embedded in the WWD provides results are comparable to or even better (i.e., faster response speed) than the laptop PVT. With the backlight on, however, the PVT response speeds tend to have larger variability on the WWD. In brief, the data in Figures 8 and 9 show that when the backlight is off, there is a systematic bias in the WWD RTs (responses fall well above the equality line). In contrast, when the backlight is on, PVT responses are better clustered around the equality line.

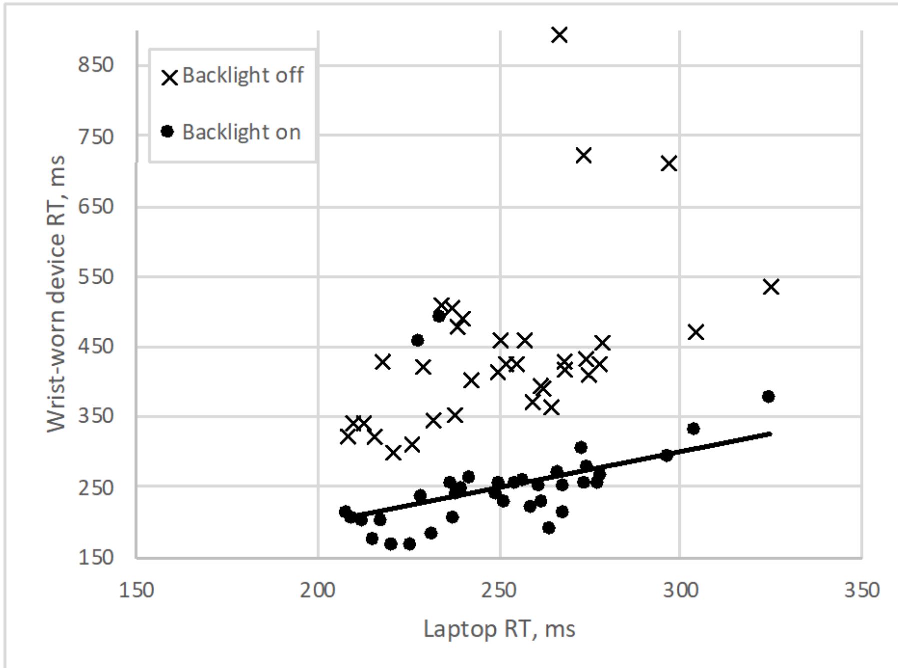


Figure 8. Reaction time in low ambient light with wrist-worn PVT backlight on (•) and off (x). The solid black line indicates the equality line when compared to the laptop PVT.

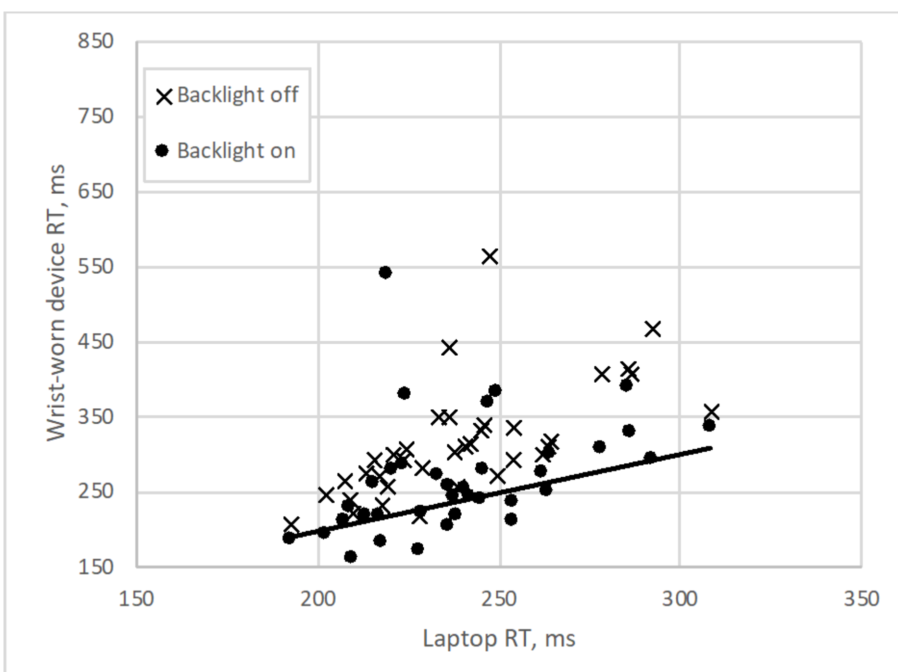


Figure 9. Reaction time in normal ambient light with wrist-worn PVT backlight on (•) and off (x). The solid black line indicates the equality line when compared to the laptop PVT.

Figures 10 to 13 show that Bland-Altman plots of the RTs in the four experimental conditions (low/normal ambient light, backlight on/off). The regression line is denoted by the solid black line. The dotted lines represent the regression-based 95% limits of agreement. It is evident that a proportional bias exists in all conditions, i.e., there is an incremental relationship between the RT differences and the magnitude of the RTs. These differences tend to be in opposing directions when the backlight is on. Specifically, negative differences (i.e., faster responses in the wrist-worn device compared to the laptop) are associated with short RTs, whereas positive differences (i.e., slower responses in the wrist-worn device compared to the laptop) are associated with long RTs. Each figure has different range in the horizontal and vertical axes for better depiction of differences.

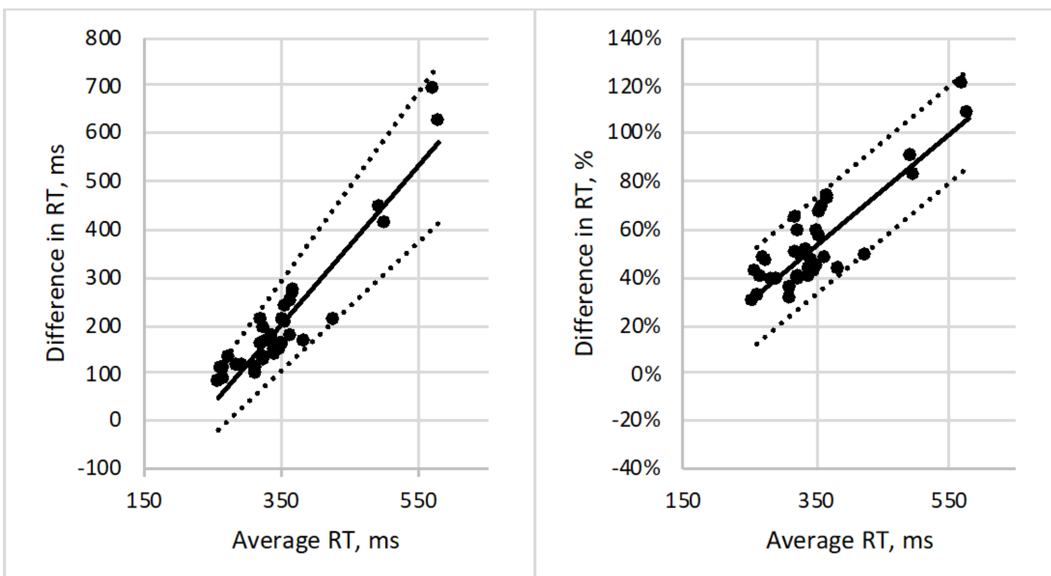


Figure 10. Absolute and percentagewise differences in PVT reaction times in low ambient light conditions. WWD with backlight off.

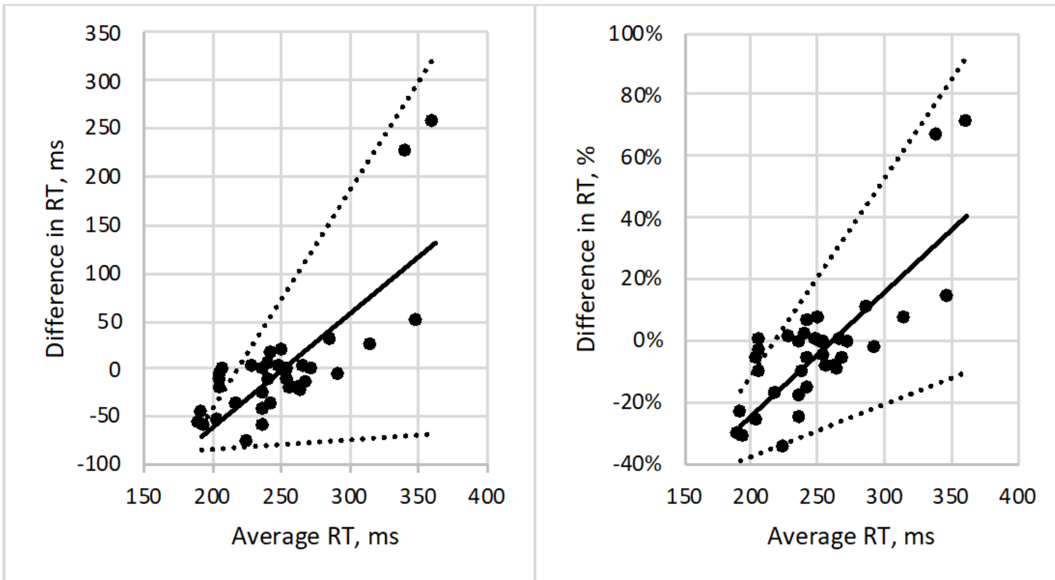


Figure 11. Absolute and percentagewise differences in PVT reaction times in low ambient light conditions. WWD with backlight on.

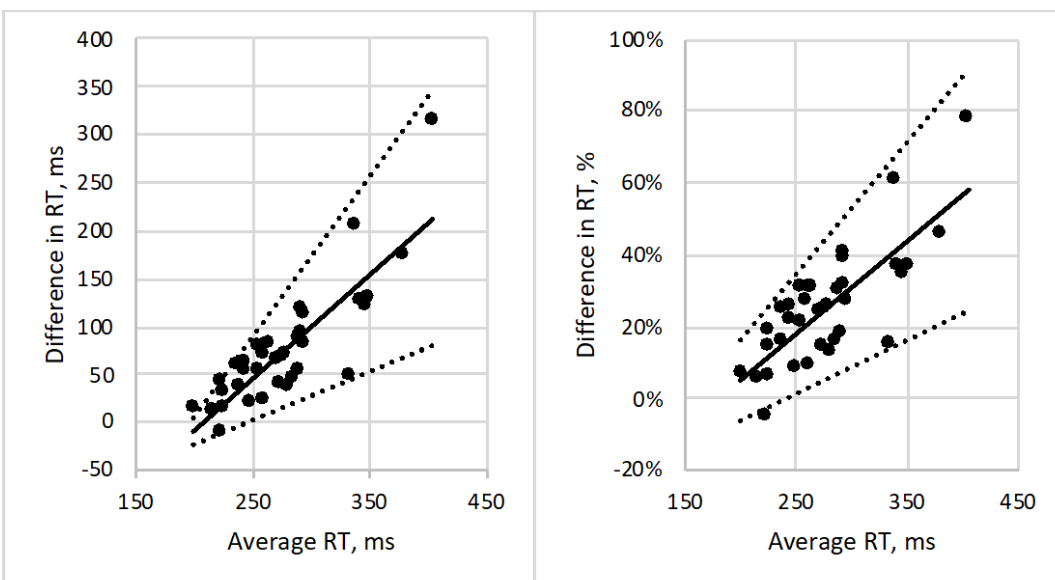


Figure 12. Absolute and percentagewise differences in PVT reaction times in normal ambient light conditions. WWD with backlight off.

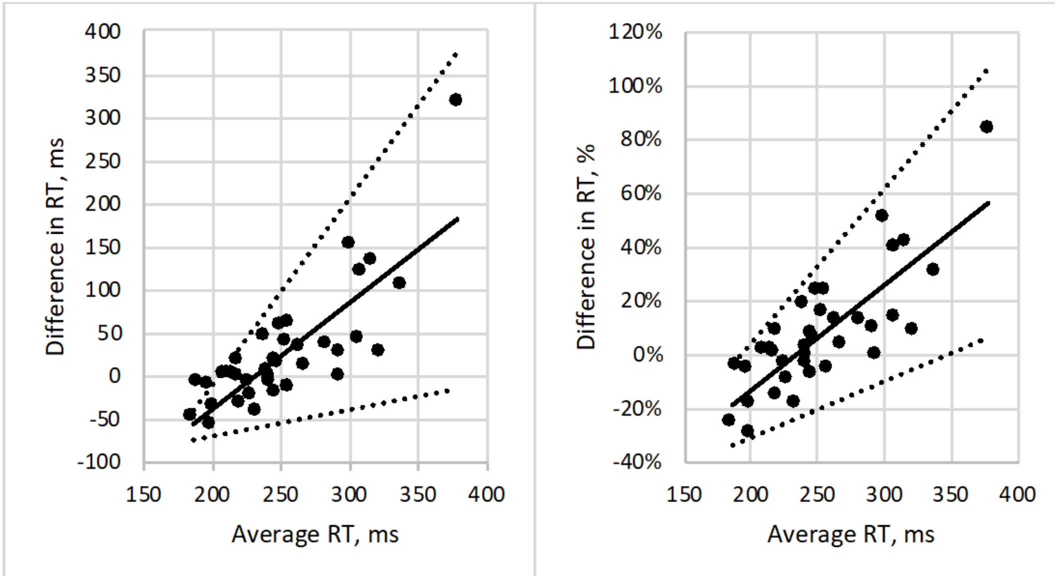


Figure 13. Absolute and percentagewise differences in PVT reaction times in normal ambient light conditions. WWD with backlight on.

The second study in this project was focused on the comparison of the 3-minute PVT (ISI = 1 – 4 seconds) embedded in a hand-held device (HHD) with a touch-screen and the WWD PVT with the validated laptop PVT. Consistent with the findings in the first experiment, WWD PVT response speeds are faster than the laptop PVT speeds when the backlight is on, but slower when the backlight is off (Figure 14) and the variability increased such that the range of values was 60% greater compared to the laptop PVT. The range of responses for the HHD PVT, on the other hand, was 60% lower than that of the laptop, and the response speeds are notably slower (Figure 15).

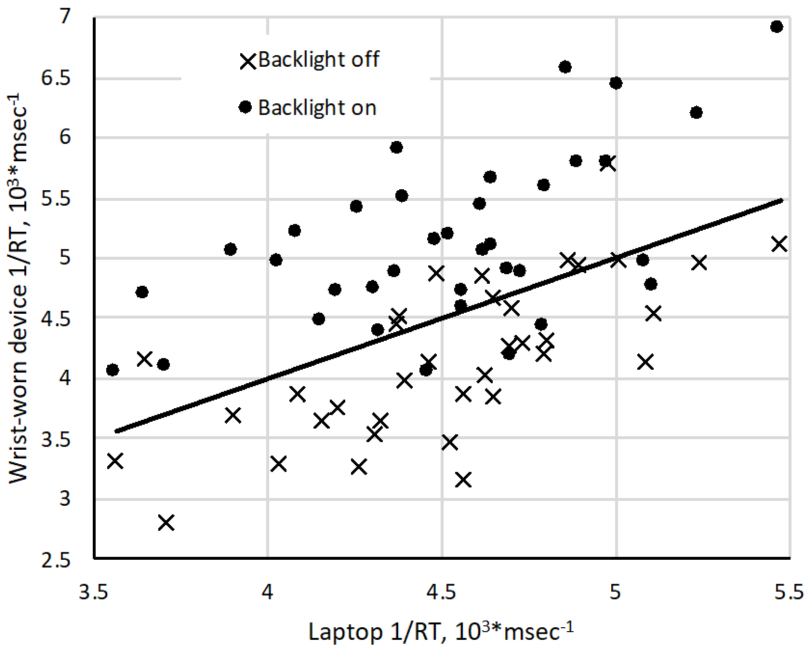


Figure 14. Response speeds with WWD PVT backlight on (•) and off (x). The solid black line indicates the equality line when compared to the laptop PVT.

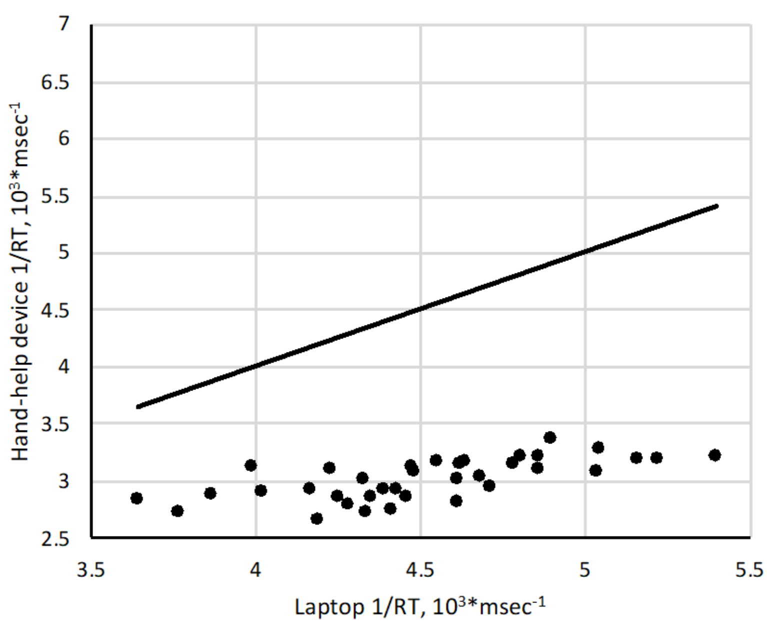


Figure 15. Response speeds with the hand-held PVT. The solid black line indicates the equality line when compared to the laptop PVT.

Next, we conducted a Bland-Altman analysis of the response speeds. Figure 16 includes six diagrams, three diagrams (A, B, C) for the absolute differences between devices and three diagrams (D, E, F) for the percentage-wise differences. The slope of the regression line between the difference (absolute and percentage-wise) and the mean of the WWD and the L PVT showed a consistent upward trend (as shown in diagrams A, B, D, and E). This incremental association between the differences in response speed and the magnitude of the average response speed was found to be in opposing directions when the backlight feature in the WWD was on. That is, fast individuals tended to perform better on the WWD than the L, whereas slow individuals tended to perform worse on the WWD than the L. Furthermore, the variability of the differences between the WWD with the backlight off and the L was constant but decreased in faster response speeds when compared to the WWD with the backlight on and the L. These trends are evident in the dotted 95% agreement limits shown in diagrams A/D and B/E, respectively.

In contrast, the slope of the regression line between the differences and the mean of the HHD and the L PVT showed a consistent downward trend (as shown in diagrams C and F of Figure 3 in conjunction with the HHD scatterplot in Figure 2) with the difference between devices constantly increasing in absolute values (i.e., the HHD seems to slow faster individuals with the magnitude of this effect increasing with faster responses).

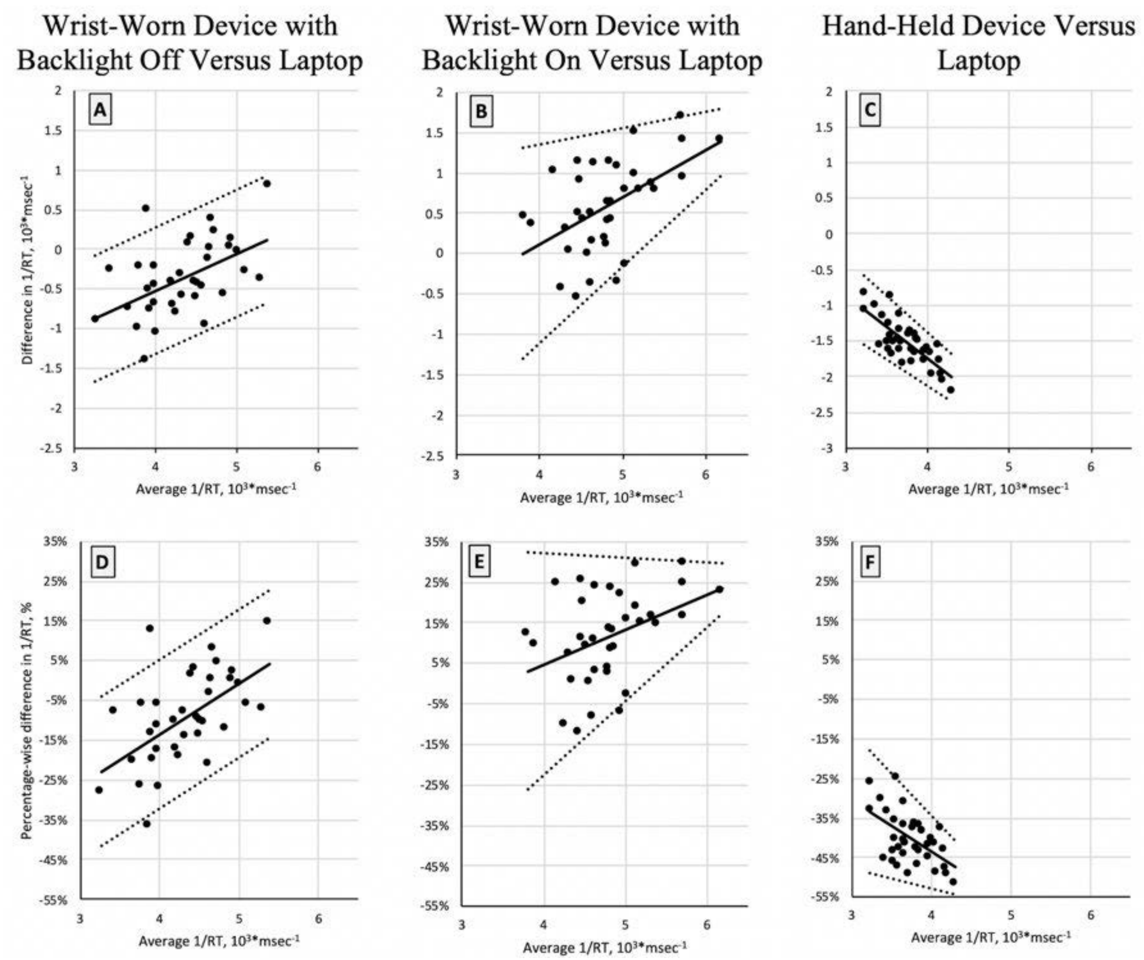


Figure 16. Bland-Altman plots for response speed. The think line represents the regression line of the absolute and percentage-wise differences. The 95% limits of agreement are represented by the dotted lines.

Given the sensitivity of PVT metrics to fatigue and circadian rhythmicity, the third study focused on the utility of the 3-minute WWD PVT to detect changes in performance between a morning and an afternoon data collection session. Also, PVT metrics from the WWD PVT were compared to the laptop PVT and Go/No-Go task. PSQI scores were used to classify participants as good sleepers (total score ≤5) or poor sleepers (total score >5). Results show that performance on both the WWD PVT and the 5-minute Go/No-Go task differ between the morning and the afternoon sessions, but the laptop PVT does not. The response speeds for the PM session are faster in both tasks compared to the AM session.

Table 5. Task metrics

Metric	AM	PM	p-value [A]
Wrist-worn PVT			
Mean RT	282±83.9	246±50.4	< 0.001 [B]
Mean Response Speed	4.18±0.72	4.50±0.73	< 0.001 [B]
False starts, %	0.65±1.81	0.33±0.79	0.253
355ms lapses, %	12.8±14.7	9.68±12.9	< 0.001 [B]
500ms lapses, %	3.97±5.10	2.67±4.66	0.119
355 ms lapses with false starts, %	13.5±14.8	10.0±13.2	< 0.001 [B]
500 ms lapses with false starts, %	4.62±6.01	3.00±5.01	0.047 [B]
Laptop PVT			
Mean RT	243±22.5	247±42.6	0.315
Mean Response Speed	4.24±0.33	4.28±0.39	0.196
False starts, %	1.72±2.23	1.76±1.70	0.897
355ms lapses, %	2.75±4.58	3.00±5.58	0.840
500ms lapses, %	0.35±0.92	0.31±0.89	0.579
355 ms lapses with false starts, %	4.47±4.94	4.76±5.74	0.970
500 ms lapses with false starts, %	2.07±2.37	2.01±2.08	0.852
Go/No-Go			
Mean RT in Go trials	311±50.4	301±44.7	0.006 [B]
Mean Response Speed in Go trials	3.39±0.45	3.50±0.44	0.001 [B]
Not responding to Go trials, %	0.11±0.31	0.22±0.45	0.202
False alarms, %	15.2±11.6	18.7±13.3	0.032 [B]

[A] Wilcoxon Signed Rank test

[B] Statistically significant based on post-hoc analysis with the BH-FDR controlling procedure

The WWD PVT results also show an association between response speed and PSQI scores, where higher PSQI scores (poor sleepers) performed worse. This association is not evident with the laptop PVT or the Go/No-Go task. Figure 17 shows the response speed by task and by the time of day when the data were collected. Of note, response speed is averaged over all PVT RTs excluding false starts, whereas the Go/No-Go task includes only responses for the Go trials.

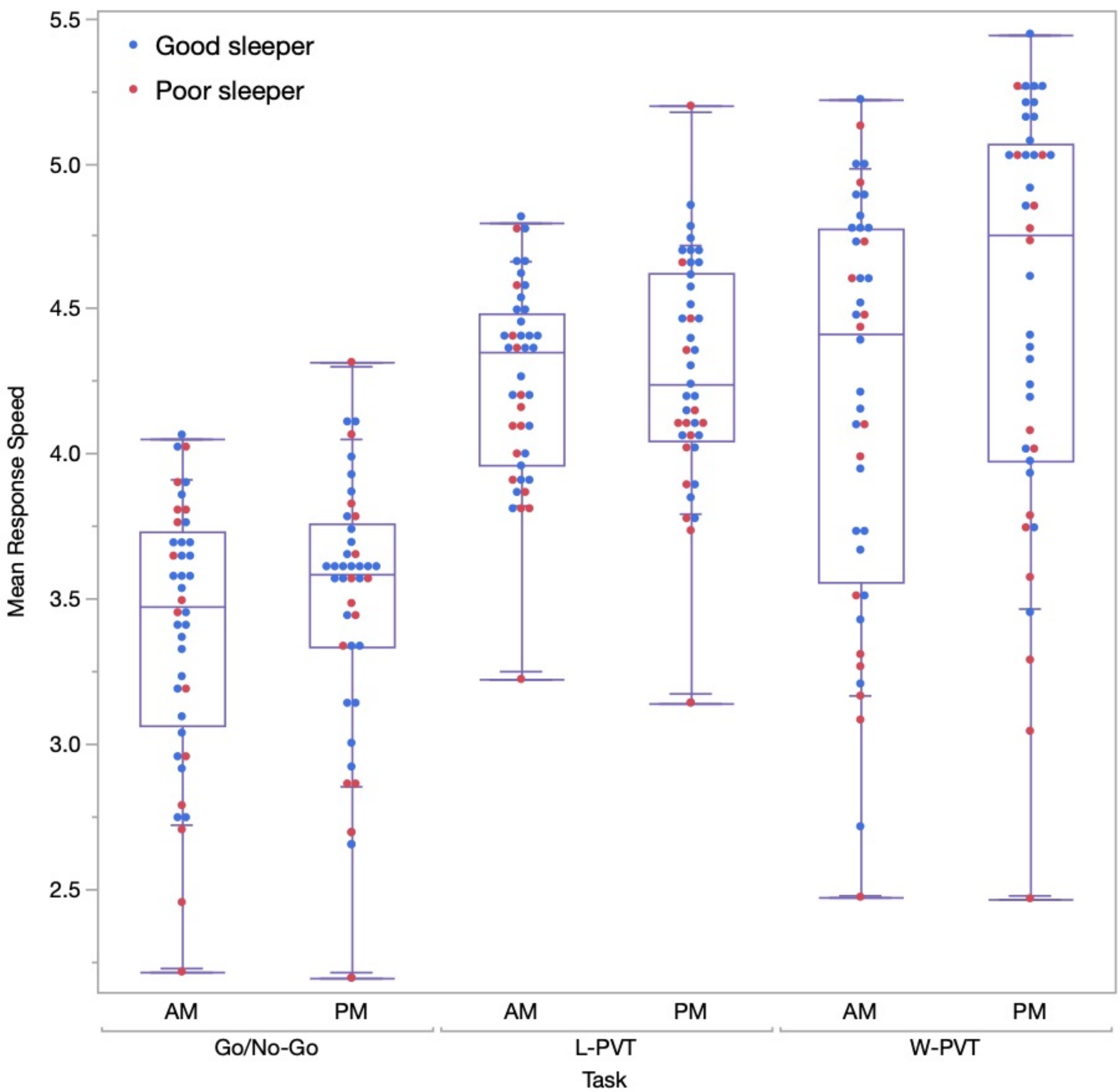


Figure 17. Quartile plots of response speed by task (Go/No-Go, laptop PVT [L-PVT], wrist-worn PVT[W-PVT]), time of day (AM, PM), and PSQI score classification.

Also, we assessed AM/PM differences in errors of commission and omission by task and time of day that the data were collected. Errors of commission included PVT false starts and responses to No-Go trials from the Go/No-Go task. We calculated the errors of omission twice. In the first calculation, they included the sum of non-responses to a target in Go trials (Go/No-Go) and the 355 millisecond lapses (PVT). In the second calculation, errors of omission included the sum of non-responses to a target in Go trials

(Go/No-Go) and the 500 millisecond lapses (PVT). Figure 18 shows errors of commission and omission by task and time of day that the data were collected.

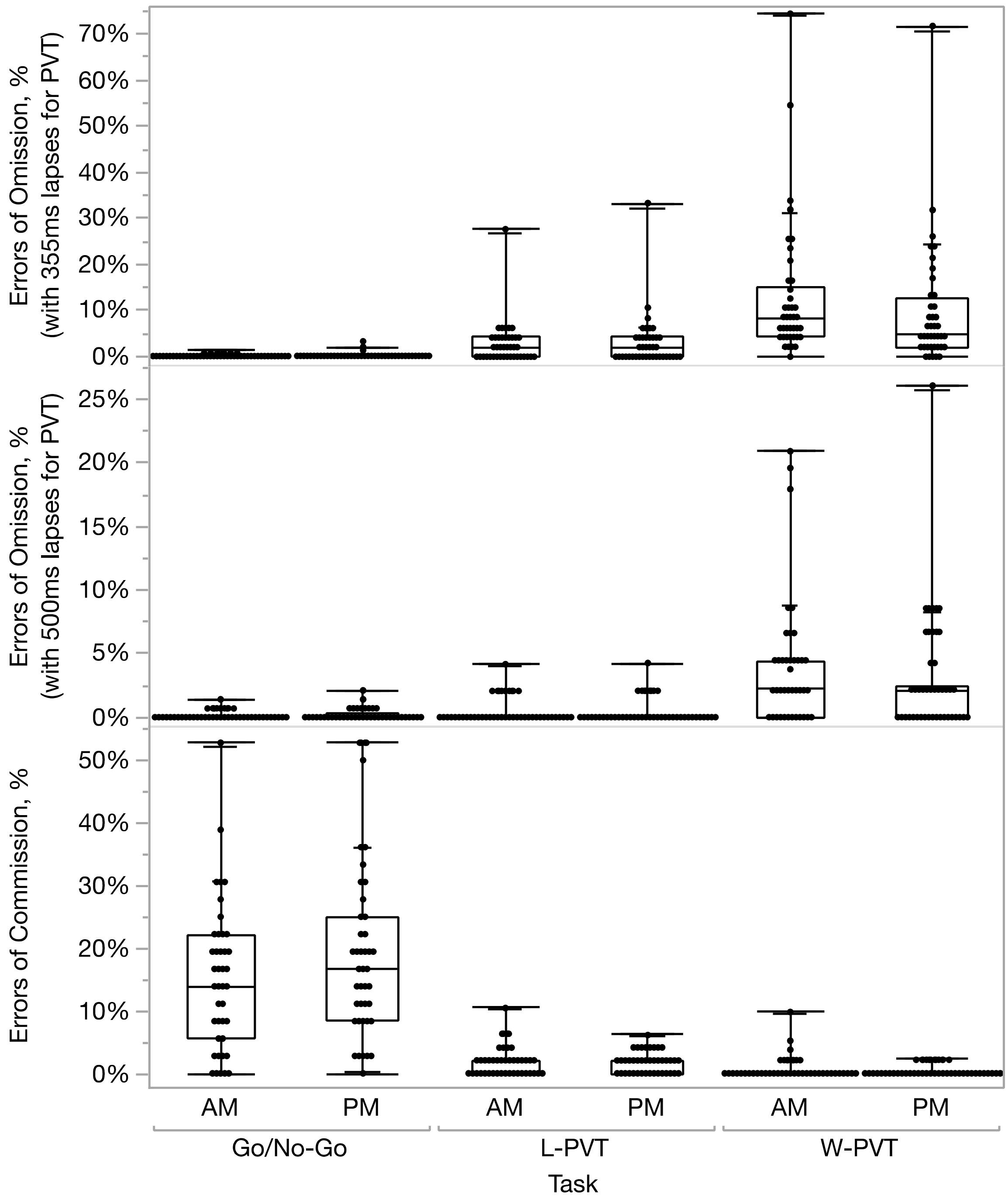


Figure 18. Quantile plots of omission and commission errors by task, time of day, and PSQI score classification (L – PVT denotes the laptop PVT; W – PVT denotes the wrist-worn PVT)

B. LESSONS LEARNED FROM COLLECTING PVT DATA IN THE FIELD

The Crew Endurance Team has been collecting PVT data in field settings for more than 15 years. In this section, we will provide a brief overview of the lessons learned in these studies in relation the difficulties in collecting and preparing PVT data for analysis, but also methods we used to overcome these difficulties.

1. The duration of PVT

The typical PVT duration is 10 minutes with an inter-stimulus interval (ISI) of 2–10 seconds. This version of PVT has been used extensively in laboratory settings, but its use in operational field studies is problematic for a variety of reasons. The first reason is the 10 minutes is a long time for performing the PVT while on duty. Anecdotal data from crewmembers on U.S. Navy ships showed that they consider the original 10-minute PVT excessively long for field studies. The second reason is that if PVT data are collected in the work area, it is almost impossible to perform the test without external disruptions. Therefore, in operational settings, the 10-minute version should be avoided in favor of the shorter, 3-minute version of the test. Research has shown that PVT versions of 3 to 5 minutes duration provide comparable results to the original 10-minute PVT (Basner et al., 2011; Basner & Rubinstein, 2011; Lamond, Dawson, & Roach, 2005; Thorne et al., 2005).

2. Identification of lapses

Typically, false starts and lapses are identified based on fixed and widely used criteria. For example, the reaction time criteria for a false start is 100 milliseconds. The identification of lapses is based on the 355 milliseconds criterion when ISI is between 2 and 5 milliseconds (which is equivalent to 1 to 4 milliseconds when we do not take into account the 1 millisecond the stimulus remains on the screen), and the 500 milliseconds criterion when ISI is between 2 and 10 milliseconds.

These criteria are fixed and do not consider the characteristics of the PVT system. A system that introduces a systematic constant bias which is substantive compared to the lapse duration criterion can lead to lapse-saturated results but also be less sensitive in identifying false starts. The problem of systematic inflation of lapses (i.e., failure to respond in a timely manner to a stimulus one is expecting) and underestimation of false starts may be ameliorated by using reaction time criteria based on the specific characteristics of the PVT system.

3. The interstimulus interval (ISI)

One of the criticisms of the PVT is the high presentation rate of stimuli. In general, vigilance tasks in the operational environment are characterized by infrequent, or even rare, occurrences of stimuli, that would be represented better by longer interstimulus intervals (ISI) (Van Wert, Horowitz, & Wolfe, 2009; Wolfe, Horowitz, & Kenner, 2005). The original 10-minute version of the PVT, as well as much of the PVT literature, focuses on ISIs of 2 to 10 seconds in length. More recently, the 3-minute version of the PVT was validated with the shorter ISI of 2 to 5 seconds. Both ISIs lead to a high presentation rate, but a larger ISI would lead to fewer response in the short duration of the PVT. Given these limitations, in all our studies we have used the ISI of 2 to 10 seconds.

4. Dedicated PVT data collection area or not?

In the first data collections the Crew Endurance conducted in field settings, participants were required to perform the PVT on a validated PVT system (to include a laptop and the PVT application) in an area dedicated for this task. This method ensured performing the PVT in controlled conditions to include absence of external interruptions. Despite its obvious benefits from a theoretical standpoint, this method did not work well in practice for a number of reasons.

First, centrally locating the testing device in a library or common access area requires participants to go out of their way to participate. Not surprisingly, study participants refused to be pulled away from their actual work to engage in an artificial task. In the frequently busy and stressful naval operational environment, asking fatigued sailors to deviate from the daily schedule was challenging. Typically, participants were instructed to perform the PVT twice per day if they did not stand watches, or before and after their shifts if they were watchstanders.

Consequently, these first studies suffered from extremely low compliance rates for performing the PVT as designed in the study protocol. That is, approximately 10% of the participating sailors complied with the PVT component of the protocol.

Another important concern was the external validity of the PVT data and whether it denoted sailor alertness level when these data were collected in a dedicated space. The artificial conditions and the requirement for sailors to go to this dedicated space before and after their shift may alter the ability of the PVT to reliably assess sailor alertness while on duty. Lastly, there was the concern of cybersecurity with some operational environments not allowing the introduction of laptops or other devices that could potentially transmit information.

Due to these concerns, we asked watchstanding sailors to perform the PVT just before and after their shift in the area where they perform their duties. This was accomplished by sailors wearing a wrist-worn actigraph which had an embedded version of the PVT (Matsangas et al., 2017). The use of actigraphs to collect PVT data led to a 6-fold improvement in sailor compliance compared to the use of the PVT laptops in dedicated spaces.

5. Compliance with the PVT protocol

As noted earlier, our typical PVT protocol for watchstanders includes two PVT trials per shift, one trial just before the commencement of the shift and one trial immediately afterwards. Participants who were non-watchstanders performed two PVT trials per day, once in the morning and once in the evening. Based on eight data collections from seven ships and 660 sailors, the overall weighted compliance rate for the PVT protocol was 60% ranging between 20% and 75% depending on the ship.

6. Number of participants in field studies with the PVT

Field studies involving the collection of PVT data should include a large number of participants. Because they are collected in un-controlled conditions, there is extra "noise" in the PVT data; therefore it is important to collect PVT data from multiple service members, with multiple trials from each participant.

7. The effect of motion and motion sickness

In the US Navy, PVT data have oftentimes been collected from service members working in moving environments. Therefore, special care is needed for time periods when the risk of motion interference is increased. In general, motion can affect PVT performance via two paths, the direct biodynamic path and the indirect through the development of motion sickness and sopite syndrome symptoms (Matsangas & McCauley, 2014; Matsangas, McCauley, & Becker, 2014). In both cases, PVT performance will deteriorate. Depending on the research question, the effect of environmental motion on PVT metrics may be an unwanted artifact to be teased out, or could also potentially be the main focus of the study.

In the Crew Endurance Team studies, our focus is on sailor well-being and the cognitive effects of various occupational stressors. Therefore, motion effects were considered a potential bias. Along these lines, our approach was to exclude time periods during which the biodynamic or motion sickness/soporific interference was expected to be elevated (e.g., rough sea conditions). Of note, however useful, this simple procedure does not fully eliminate the problem of carry-over effects.

8. Preparation of field PVT data

Approximately 70% percent of the effort in analyzing PVT data collected in the field is preparing them for analysis. Given that these data have been collected without a researcher in immediate attendance, multiple layers of scrutiny are needed in order to ensure a minimum level of quality before adding these data to the overall datafile for further analysis.

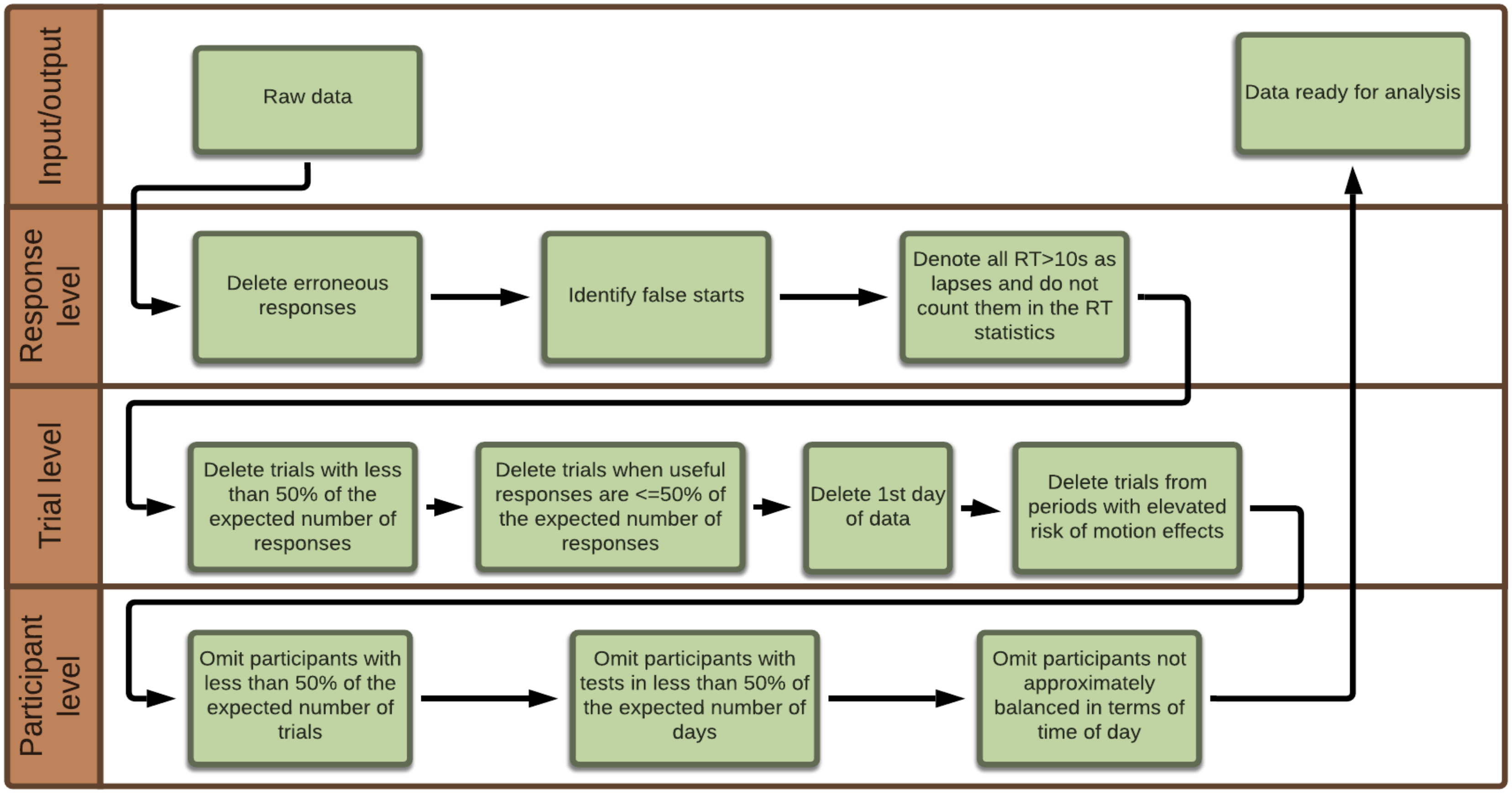
The procedure we have developed and refined over the years has three main levels each of which is focusing on the PVT data. The response level is focused on scrubbing the raw PVT responses. The first step is to omit from further analysis any pseudo-responses, i.e., stimuli which have not been responded by the user. Typically, these non-responses are identified by a large reaction time (e.g., 65535 milliseconds) and may denote a time when the user has been distracted from performing the PVT due to extraneous reasons. Next, false starts and lapses must be identified and classified as such.

The next level focuses on the trial itself. Responses which have been scrubbed are aggregated by trial. The first step is to omit from further analysis PVT trials with less than 50% of the expected number of responses in each trial. Next, trials with fewer than 50% of useful responses (responses with reaction times to be used for the calculation of reaction time/response speed metrics – false starts do not contribute to this calculation). We also omit any trials performed on the first day of the data collection. Oftentimes trials on the first day have multiple false starts and pseudo-responses as participants are getting familiar with the device. Lastly, we omit any trials which were performed during periods where the risk of motion effects is elevated, e.g., high sea state may increase the risk of having biodynamic artifacts in the PVT responses or having participants who are motion sick.

The last level of scrutiny focuses on PVT data by participant. Cleaning data at this level will identify compliance with the research protocol and lead to the decision of which participants will be included for the analysis of the PVT data. First, participants with fewer than 50% of the expected number of trials should be omitted from further analysis. Also, omit participants with trials on fewer than 50% of the data collection days. Lastly, omit participants with trials which are not approximately balanced in terms of time of day. For example, some sailors may perform the PVT only before or after their shift, or only during morning but not nights, etc.

Our method ensures that PVT data are representative of the entire data collection period, approximately evenly balanced throughout the data collection period and within each day. Of note, however, our method should be followed by outlier identification to further ensure that PVT data are ready for analysis. Our method is shown in the following figure.

Figure 19. Diagrammatic depiction of PVT data scrubbing procedure



IV. CONCLUSIONS AND RECOMMENDATIONS

This section provides an overview of the project and discusses our findings in terms of the three aims of the project.

AIM 1. VALIDATE THE 3-MINUTE PVT EMBEDDED IN ACTIGRAPHS

The results of our experiments conducted under this project demonstrate that when the screen backlight is on, the PVT embedded in the AMI actigraphs provides results comparable to the laptop PVT. Therefore, the 3-minute PVT in the AMI actigraphs is a valid and alternative to the 3-minute laptop-based PVT for field assessment. We believe that the backlight feature is important when performing the PVT on the AMI Motionlogger because the device has a low contrast LCD screen. PVT performance in newer devices with high contrast screens will not be as sensitive to ambient lighting conditions.

More generally, findings from our studies in controlled conditions, combined with lessons learned from our field studies, clearly demonstrate that having the PVT embedded in a wrist-worn device is a useful method to collect PVT data in the operational environment. The AMI PVT provides a simple, less disruptive, and rapid operational test of psychomotor vigilance performance which could be useful for occupations outside the military, such as security-related occupations, first responders and emergency medical teams, power grid and plant operators doing shift work, and air traffic controllers.

Our findings also shed light on an issue of concern, that is, the comparison of PVT metrics between different devices or the same devices using different settings. The results of our research effort underscore the importance of avoiding direct comparison of PVT performance metrics when using different devices or same devices with different settings. In such cases, researchers should expect systematic differences in terms of constant bias, variability, and/or transformation of reaction times.

AIM 2. EXPLORE THE UTILITY OF OTHER DEVICES FOR FIELD USE

Given the widespread use of hand-held devices with a touch screen interface, a large number of studies have investigated the use of tablets and other portable devices to collect PVT data. Overall, the findings of these studies suggest that the PVT embedded in touch screen devices is sensitive to wakefulness and, therefore, such systems could be used for measuring the effects of fatigue. Based on this background information, we tested a touch screen device with a PVT application. This representative PVT system had disappointing results. That is, the hand-held device using a touch screen interface to collect PVT data introduced a large constant bias and, more importantly, a proportional bias that decreased the range of response speeds compared to the validated PVT.

These findings raised the issue of what the appropriate interface should be for a field-grade PVT system, especially if we consider that many service members are working on moving platforms. In our study, participants were asked to hover their finger a small distance from the screen while waiting for the stimulus. Such methods should be avoided in the naval operational environment, in combat vehicles, or aircrafts. Pressing a button is more appropriate in field settings, but not all devices have buttons. Furthermore, the mechanical characteristics of button may be an issue of concern in field settings, e.g., sensitive buttons may not be appropriate. In general, given that many researchers are using touch screen devices for collecting PVT data, future efforts should assess the effect of environmental motion on PVT metrics and determine the optimal interface in order to collect reliable PVT data in the field.

AIM 3. PROVIDE RECOMMENDATIONS FOR THE COLLECTION OF PVT DATA IN THE MILITARY OPERATIONAL ENVIRONMENTS

The Psychomotor Vigilance Task (PVT) has existed for more than three decades and is validated on multiple devices with different parameters. It is widely accepted in the sleep and fatigue community as a method to assess somnolence and fatigue effects. The task itself is easy to perform, does not require special and lengthy training, does not show any learning effects, and the 3-minute version is appropriate for operational use.

The utility of the PVT application in operational environments, however, should not be distinguished from the device in which it is embedded. The PVT system includes both the software and the hardware, which must be validated and assessed for its utility in field settings. Based on our 10+ year experience collecting PVT data in the various military environments, combined with our findings in this project, we believe the PVT system is optimal in field settings if the PVT is embedded in a wearable device. Given that the wearable device would typically replace a watch, it should incorporate the functionality of a typical watch combined with the specific characteristics appropriate for the PVT. That is, it should have a high contrast screen to allow for changes in ambient light, and a mechanical button which allow for more reliable PVT responses. Touch screens should be avoided as an interface to collect PVT data.

Portable devices (like smartphones and tablets) may be useful as well. Actually, most of the recent PVT literature focuses on such devices. Useful as they may be, however, portable devices may pose an extra burden for the service members given that these devices need to be carried if data are to be collected in field conditions. Also, the touch screen interface is a concern if PVT data are collected in moving environments.

Based on our experience, two issues are common in PVT field studies: missing data and artifacts due to extraneous factors (i.e., not related to the performer’s state). The method we propose reduces bias due to missing data and artifacts from external disturbances by assessing the quality of the PVT data at three levels, the raw response level, the aggregated (trial) level, and the participant level.

Lastly, we explored the availability of wrist-worn devices which can be used for activity/sleep assessment and include an embedded version of the PVT. Unfortunately, other than the AMI Motionlogger that has been discontinued, we failed to identify any such devices during this project. However, we believe that it will be valuable to have an integrated wrist-worn device to include the functions of a watch, a research grade actigraph, and a device to collect PVT data.

APPENDIX A: LIST OF PRODUCTS DEVELOPED UNDER THIS PROJECT

A. REFEREED JOURNALS ARTICLES

Matsangas, P. and Shattuck, N.L. (2020). Hand-held and wrist-worn field-based PVT devices versus the standardized laptop PVT. *Aerospace Medicine and Human Performance*, 91(5), 409-415, doi:10.3357/AMHP.5567.2020.

Matsangas, P., Shattuck, N.L., and Brown, S. (2017). Preliminary validation study of the 3-minute wrist-worn psychomotor vigilance test. *Behavioral Research Methods*, 49(5), 1792-1801, doi:10.3758/s13428-016-0821-2.

B. REFEREED CONFERENCE AND SYMPOSIA PAPERS

Matsangas, P., Shattuck, N.L., Mortimore, K., Paghasian, C., and Greene, F. (2019). The 3-minute Psychomotor Vigilance Task (PVT) Embedded in a Wrist-worn Device: Time of Day Effects. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 63(1), 797-801, doi:10.1177/1071181319631144.

Matsangas, P., and Shattuck, N.L. (2018). Agreement between the 3-minute Psychomotor Vigilance Task (PVT) Embedded in a Wrist-worn Device and the Laptop-based PVT. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 62(1), 666-670, doi:10.1177/1541931218621151.

C. CONFERENCE POSTERS

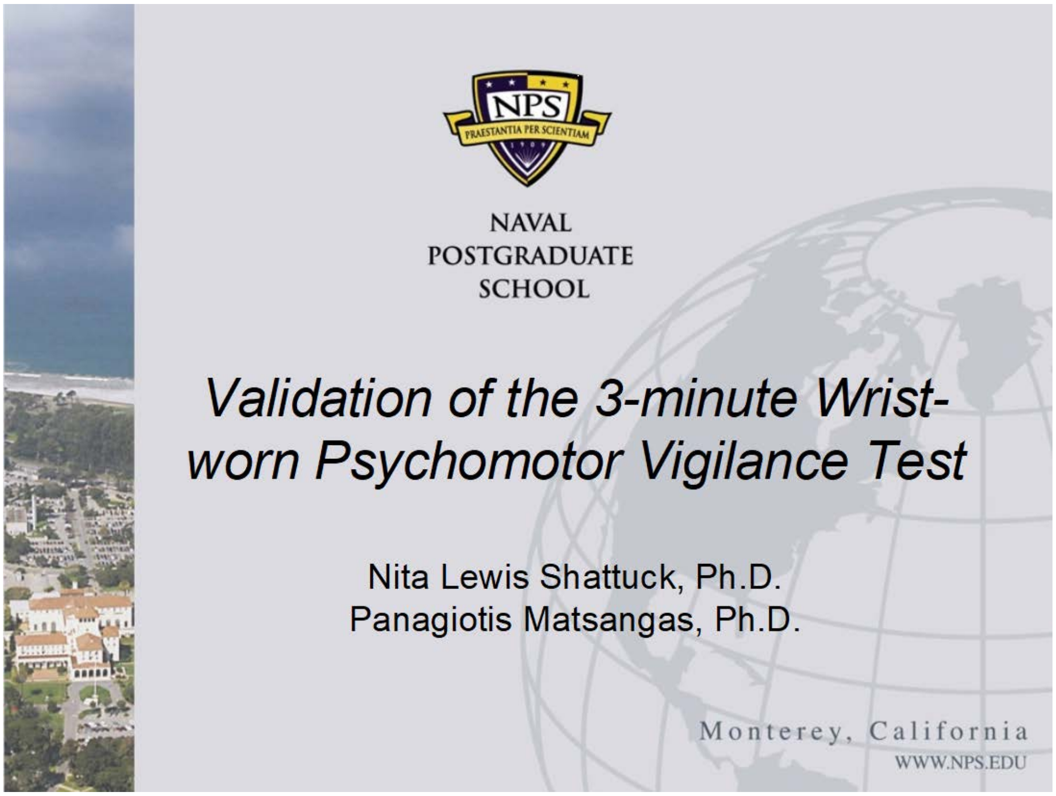
Shattuck N.L., Matsangas P., and Sjörs Dahlman, A. (2017, 27-30 August). Validation of the Psychomotor Vigilance Task (PVT) in the Operational Environment: Exploring the Utility of Different Devices and the Effect of Environmental Factors. Poster presented at the Military Health System Symposium (MHSRS), Kissimmee, FL.

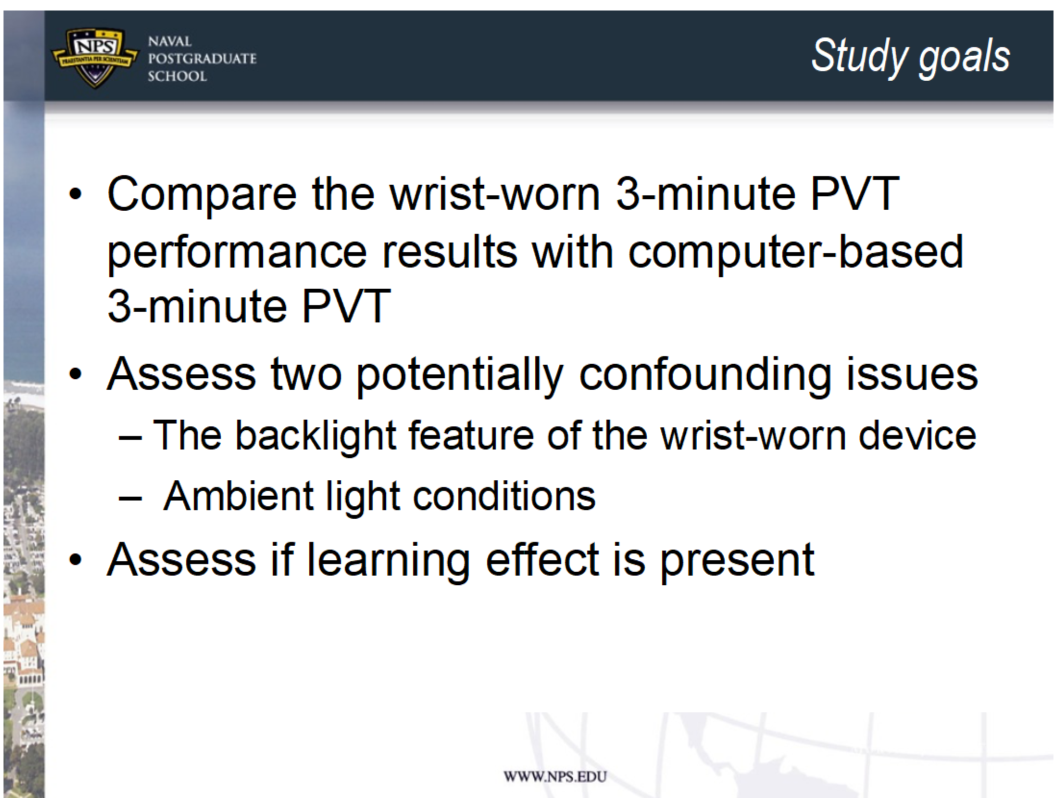
D. PRESENTATIONS

Presentation with the results of FY2015.

Presentation with the interim results of FY2016.

APPENDIX B: PRESENTATION WITH THE FY2015 RESULTS





NAVAL POSTGRADUATE SCHOOL

Study overview

- Design
  - Repeated-measures experiment with three factors:
    - PVT device type (laptop, wrist-worn actiwatch)
    - Red backlight feature of the wrist-worn device (on, off)
    - Ambient light conditions
      - Low: similar to twilight (2 – 3 lux)
      - Normal: office lighting environment (300 – 400 lux)
- Setting
  - Controlled laboratory environment (NPS HSI Lab)
- Participants
  - Thirty-six healthy subjects (eight females)
  - Average age: 36.0±7.47 years old

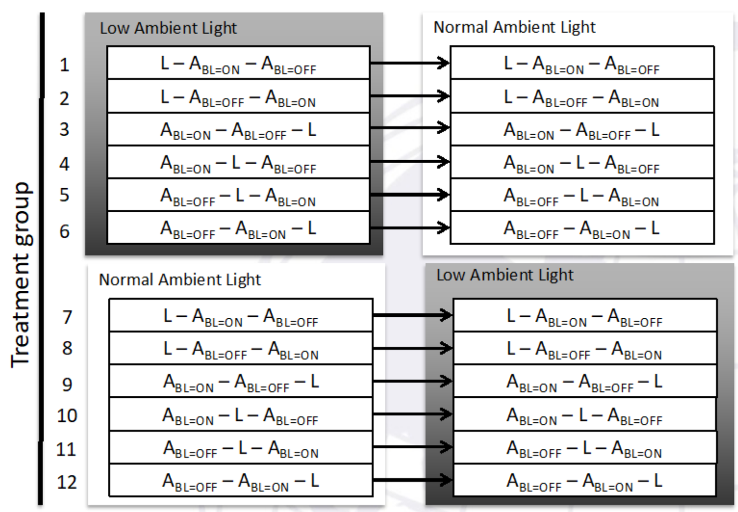
WWW.NPS.EDU

NAVAL POSTGRADUATE SCHOOL

Experimental design: 12 treatment groups (3 per group)

<u>Legend</u>

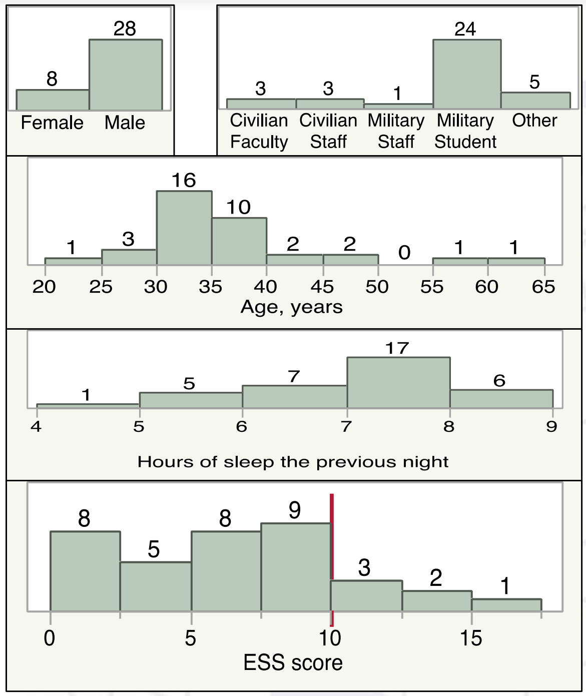
- L: Laptop
- A: Actiwatch
- BL: Actiwatch Backlight



WWW.NPS.EDU

Study overview Participants

- Thirty-six healthy adults from varying occupations
- Average age: 36.0±7.47 yr
- Reported sleep on previous night: 6.75±1.0 hr
- Alertness: Epworth Sleepiness Scale (ESS) avg = 6.19±4.0
  - 5 participants with elevated daytime sleepiness (ESS>=10)



Phase 1 Preliminary Findings

- When the wrist-worn PVT backlight is on
  - Difference in PVT metrics between the laptop and the wrist-worn versions in both ambient light conditions are not substantive, i.e., less than 8%
- Wrist-worn PVT performance with the backlight on
  - Is relatively insensitive to changes in ambient light conditions
- In normal lighting conditions
  - No difference in performance for backlight on or off for the wrist-worn PVT

WWW.NPS.EDU

NAVAL POSTGRADUATE SCHOOL

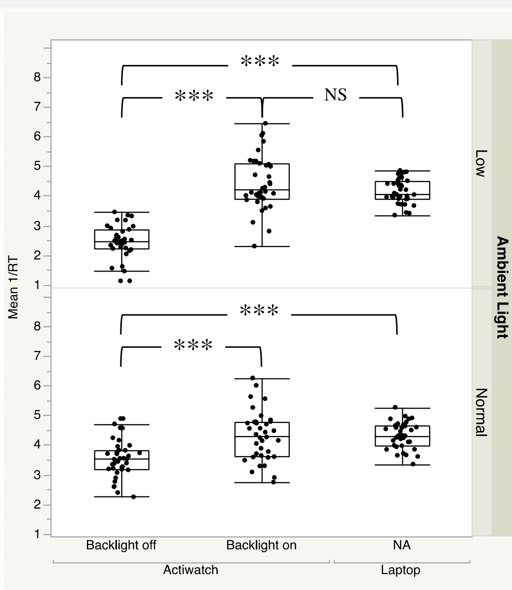
Results

- When **backlight is on** the wrist-worn PVT
  - Median difference in PVT metrics ranges between -8% (for the fastest 10% RTs) to 4.55% (for the mean 1/RT)
  - No significant difference identified for median RT and the percentage of 355ms lapses combined with false starts.
  - Wrist-worn PVT performance when the backlight is on is relatively sensitive to changes in ambient light conditions
- When **backlight is off** on wrist-worn PVT
  - Compared to the laptop, the wrist-worn PVT metrics significantly deteriorated
  - In **low ambient light** conditions, the wrist-worn shows
    - 60% increase in mean/median/fastest 10% RTs
    - % of 355ms lapses, % of 355 ms lapses combined with false starts: 77% increase
    - % of 500ms lapses, % of 500 ms lapses combined with false starts: 11% to 13% increase
  - In **normal ambient light**, the wrist-worn shows
    - 22% to 28% increase in mean/median/fastest 10% RTs
    - % of 355ms lapses, % of 355 ms lapses combined with false starts: 8% to 10% increase
- No statistically significant learning effects—asymptotic performance reached **after one trial** with wrist-worn PVT

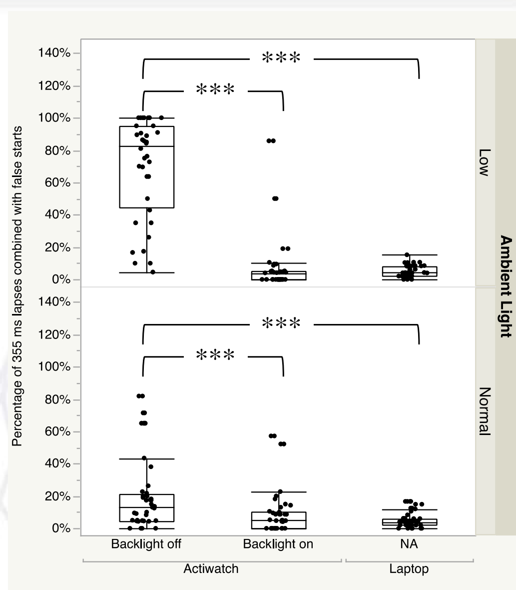
WWW.NPS.EDU

NAVAL POSTGRADUATE SCHOOL

Results

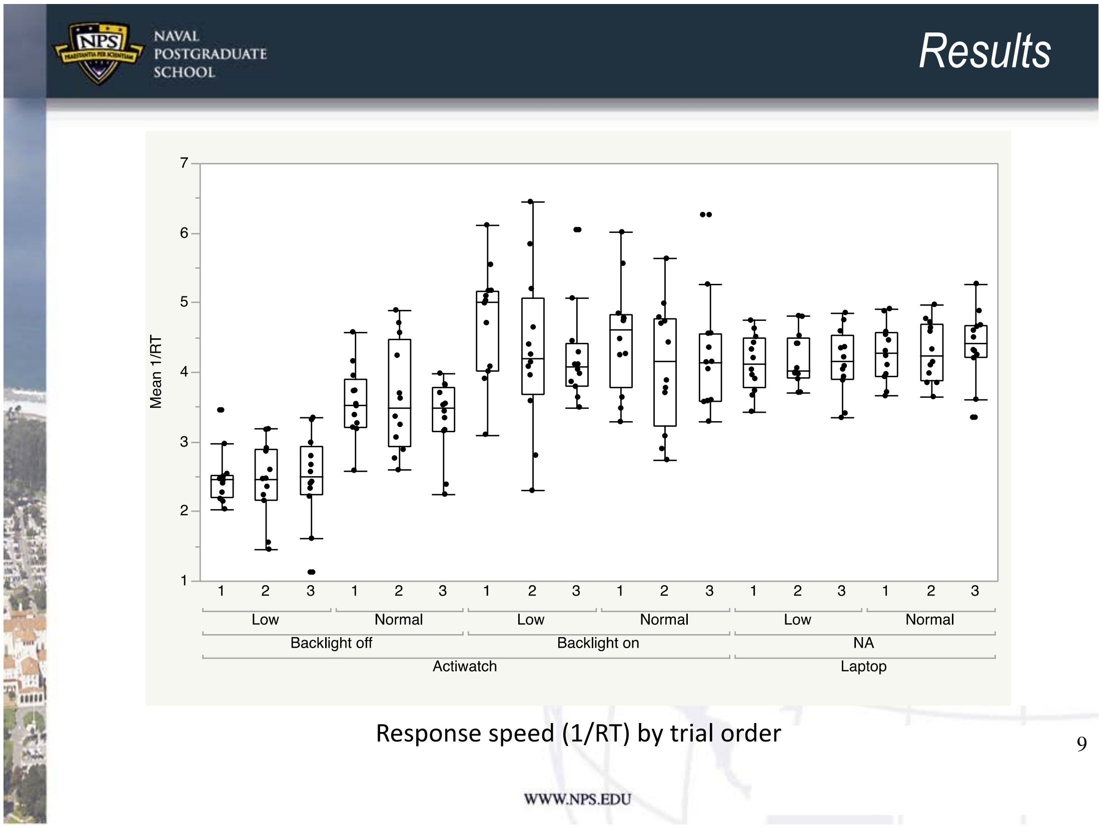


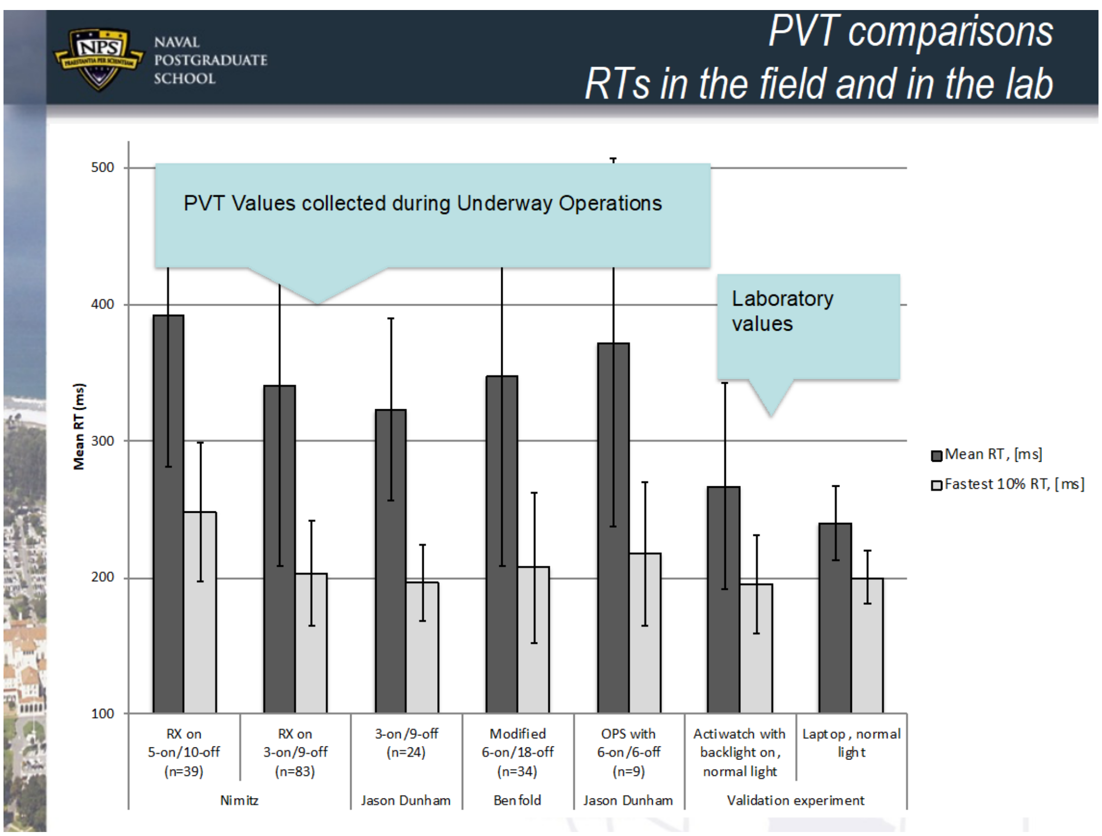
Response speed (1/RT)



Percentage of 355 ms lapses combined with false starts

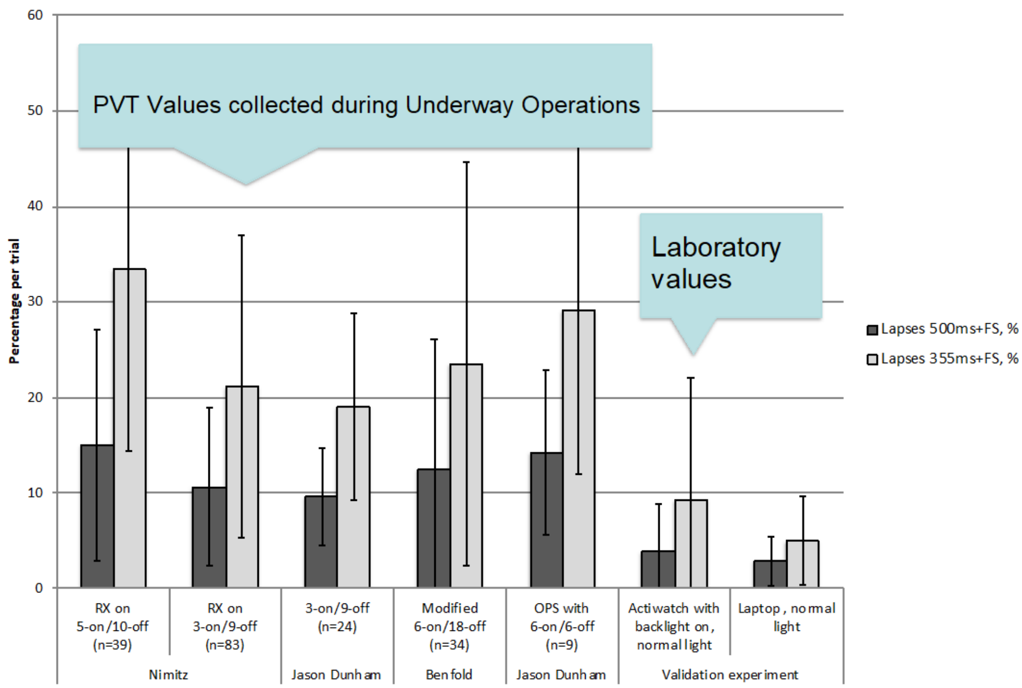
WWW.NPS.EDU


NAVAL
POSTGRADUATE
SCHOOL
Results
Mean 1/RT
1 2 3 4 5 6 7
1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3
Low Normal Low Normal Low Normal
Backlight off Backlight on NA
Actiwatch Laptop
Response speed (1/RT) by trial order
9
WWW.NPS.EDU


NAVAL
POSTGRADUATE
SCHOOL
PVT comparisons
RTs in the field and in the lab
PVT Values collected during Underway Operations
Laboratory values
Mean RT (ms)
100 200 300 400 500
Mean RT, [ms]
Fastest 10% RT, [ms]
RX on 5-on/10-off (n=39)
RX on 3-on/9-off (n=83)
3-on/9-off (n=24)
Modified 6-on/18-off (n=34)
OPS with 6-on/6-off (n=9)
Actiwatch with backlight on, normal light
Laptop, normal light
Nimitz
Jason Dunham
Benfold
Jason Dunham
Validation experiment

NAVAL POSTGRADUATE SCHOOL

PVT comparisons
Errors (Lapses+FS) in the field and in the lab



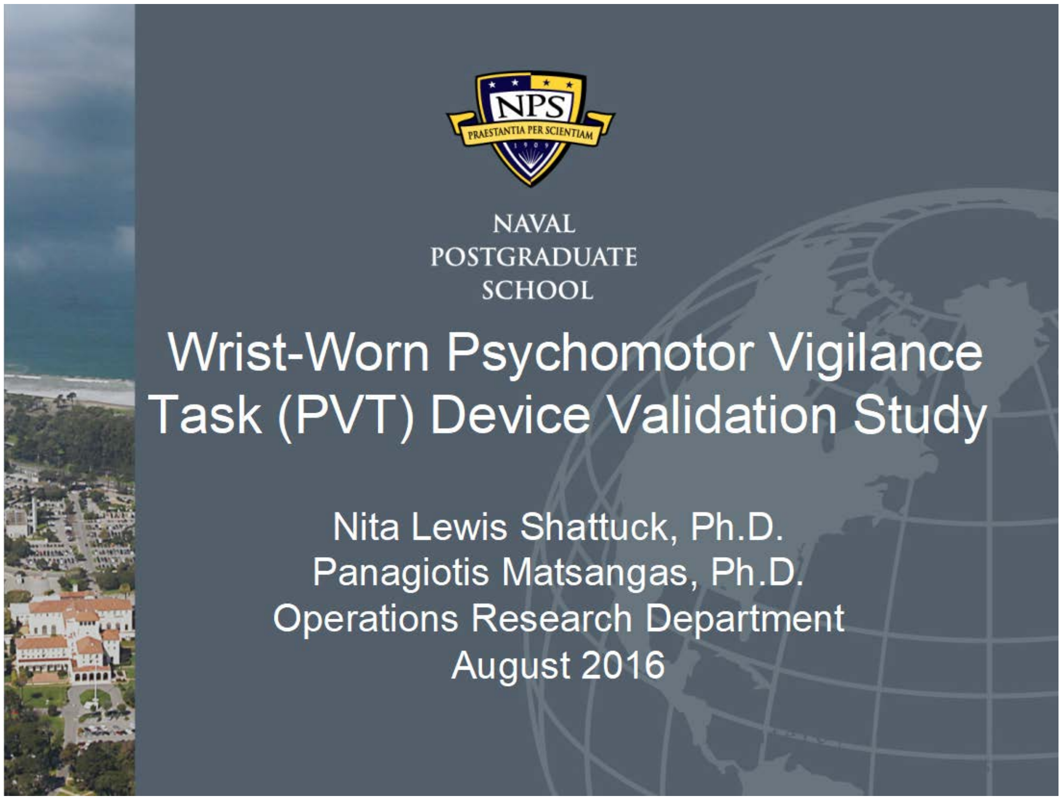
NAVAL POSTGRADUATE SCHOOL

Conclusions

- Accomplishments:
  - Initial comparison of AMI Wrist-worn PVT and Pulsar laptop version of PVT
  - Validated use of the wrist-worn PVT in past NPS field studies
- Way forward
  - Replicate study with different PVT ISIs
  - Compare with WRAIR's PVT; Apple, Android apps?
  - Make recommendations for use of PVTs in the field

WWW.NPS.EDU

APPENDIX C: PRESENTATION WITH THE INTERIM RESULTS OF FY2016



NAVAL POSTGRADUATE SCHOOL

PVT

- Objectives
  - Compare the wrist-worn 3-minute PVT results with the laptop-based 3-minute PVT when the ISI for both devices is 1 to 4 seconds
  - Assess the effect of the backlight feature provided in the wrist-worn device when ISI is 1 to 4 seconds
  - Determine the magnitude of the learning effect for the 3-minute wrist-worn PVT when ISI is 1 to 4 seconds
  - Assess the feasibility of DoD-wide individual use for fitness-for-duty applications

2

WWW.NPS.EDU

NAVAL POSTGRADUATE SCHOOL

Study overview

- Design
  - Repeated-measures experiment with two factors:
    - PVT device type (laptop, wrist-worn actiwatch)
    - Red backlight feature of the wrist-worn device (on, off)
  - Ambient light conditions
    - Normal: office lighting environment (300 – 400 lux)
- Setting
  - Controlled laboratory environment (NPS HSI Lab)
- Participants
  - Sixty-three healthy subjects

WWW.NPS.EDU

NAVAL POSTGRADUATE SCHOOL

Experimental design: 6 treatment groups

Day 1

- Recruitment
- Sign consent form
- Complete Study questionnaire
- Receive actiwatch and sleep log

Random Assignment →

Day 3

Normal ambient light

Treatment group	
1	L – $A_{BL=ON}$ – $A_{BL=OFF}$
2	L – $A_{BL=OFF}$ – $A_{BL=ON}$
3	$A_{BL=ON}$ – $A_{BL=OFF}$ – L
4	$A_{BL=ON}$ – L – $A_{BL=OFF}$
5	$A_{BL=OFF}$ – L – $A_{BL=ON}$
6	$A_{BL=OFF}$ – $A_{BL=ON}$ – L

Legend

- **L**: Laptop
- $A_{BL=ON}$: Actiwatch with the backlight ON
- $A_{BL=OFF}$: Actiwatch with the backlight ON

WWW.NPS.EDU

NAVAL POSTGRADUATE SCHOOL

PVT Tasks completed

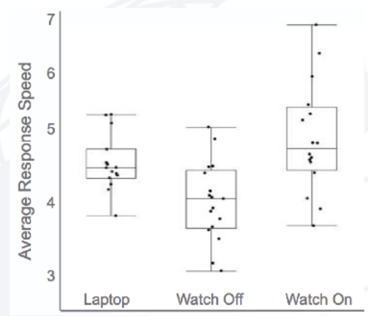
- Collected data with the wrist-worn PVT and compared with the validated laptop PVT
  - Both devices with ISI=1-4s
  - Two data collections
    - n = 39
    - n = 19
- Preliminary findings
  - The wrist-worn device with the backlight on provides results comparable to the laptop
  - In extremely low ambient light conditions, differences are evident but small

5

WWW.NPS.EDU

PVT
Preliminary findings

- Wrist-worn device
  - With backlight off
    - NOT recommended for use in operational settings
  - With backlight on
    - Average results comparable to the laptop
    - Increased variability (problem ameliorated with a large sample of data points)
    - Recommended for us in operational settings
- In extremely low ambient light conditions, differences may be more pronounced but small



- Ongoing analysis

PVT
Tasks completed

- Pilot study with tablet PVT
- Results to inform our main data collection with the tablet and compare to the laptop PVT in FY 2017

APPENDIX D: ABSTRACT ACCEPTED TO THE MILITARY HEALTH SYSTEM SYMPOSIUM (MHSRS) 2017

Title: Validation of the Psychomotor Vigilance Task (PVT) in the Operational Environment: Exploring the Utility of Different Devices and the Effect of Environmental Factors

Authors: Nita Lewis Shattuck, Panagiotis Matsangas, and Anna Sjörs Dahlman
Operations Research Department, Naval Postgraduate School, Monterey, California

Background
In military operations, it is vital to understand and quantify how individual human performance, especially vigilance and attention, varies as a function of factors such as stress and fatigue. To ensure external validity, measurements should be made during actual field operations rather than only laboratory settings. One computer-based task used extensively in controlled laboratory settings to assess the vigilance aspect of cognitive performance is the Psychomotor Vigilance Test (PVT). The utility of the PVT to capture reliable field data is affected not only by the task per se, but also by the fielding method, that is, the device used to collect data, the environmental conditions, task invasiveness, etc. Therefore, finding validated systems (i.e., field-worthy devices with validated tests) that can capture reliable and valid human performance continues to be a challenge. This multi-year project assesses the utility of wrist-worn and hand-held devices to collect reliable data with the 3-minute PVT as compared to the standardized laptop-based PVT.

Methods
We conducted three experiments ($N=91$) in laboratory conditions. The first experiment assessed how PVT performance was affected by device type (laptop/wrist-worn), backlight feature of the wrist-worn device (backlight on/off), ambient light (low/normal), and inter-stimulus interval (wrist-worn ISI=2-10 seconds, laptop ISI=1-4 seconds). The second experiment compared the wrist-worn PVT with the laptop-based PVT in normal ambient light (for both devices, ISI=1-4 seconds). The third experiment compared the PVT embedded in a touch-screen device with the laptop-based PVT.

Results
When the backlight was on, the difference in PVT metrics between the wrist-worn device and the laptop PVT was small (median $\Delta=10$ ms, $p<0.05$) or not statistically significant (combined 350 ms lapses and false starts, $p>0.5$). The touch-screen device, however, had significant average differences in PVT metrics compared to the laptop (reaction time: median $\Delta=115$ ms, $p<0.001$; combined 350 ms lapses and false starts, median $\Delta=23.3\%$, $p<0.00$). Variability in the wrist-worn and touch-screen devices was greater than the laptop.

Conclusions
Results suggest that the 3-minute PVT (ISI=2-10 seconds) embedded in the wrist-worn device can be used to collect field data. To alleviate the problem of inconsistent ambient

light conditions and increased variability, researchers should have the backlight on and increase the number of participants used in their field studies. The touch-screen device is substantively different from the laptop-based PVT both in terms of mean differences and variability. These preliminary results suggest that the use of touch-screen devices may be inappropriate for the collection of PVT data in field settings. This problem may be further exacerbated when collecting PVT in moving environments, such as ships, combat vehicles and aircraft. Future efforts will further assess the effect of environmental factors on collection of reliable PVT data in field settings using hand-held devices.

Support
These studies were supported by the Naval Medical Research Center Advanced Medical Development Program.

Disclaimer
The views expressed in this study are those of the authors and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, or the U.S. Government.

LIST OF REFERENCES

Altman, D. G., & Bland, J. M. (1983). Measurement in medicine: The analysis of method comparison studies. *Statistician*, *32*(3), 307-317.

Ancoli-Israel, S., Cole, R., Alessi, G., Chambers, M., Moorcroft, W., & Pollak, C. P. (2003). The role of actigraphy in the study of sleep and circadian rhythms. *Sleep*, *26*(3), 342-392.

Arsintescu, L., Kato, K. H., Cravalho, P. F., Feick, N. H., Stone, L. S., & Flynn-Evans, E. E. (2019). Validation of a touchscreen psychomotor vigilance task. *Accident Analysis and Prevention*, *126*, 173-176. doi: doi.org/10.1016/j.aap.2017.11.041

Arsintescu, L., Mulligan, J. B., & Flynn-Evans, E. E. (2018). Evaluation of a psychomotor vigilance task for touch screen devices. *Human Factors*, *59*(4), 661-670.

Basner, M., & Dinges, D. F. (2011). Maximizing sensitivity of the Psychomotor Vigilance Test (PVT) to sleep loss. *Sleep*, *34*(5), 581–591.

Basner, M., Hermosillo, E., Nasrini, J., McGuire, S., Saxena, R., Moore, T. M., . . . Dinges, D. F. (2018). Repeated administration effects on psychomotor vigilance test performance. *Sleep*, *41*(1). doi: 10.1093/sleep/zsx187

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*(11-12), 949-959.

Basner, M., & Rubinstein, J. (2011). Fitness for duty: a 3-minute version of the Psychomotor Vigilance Test predicts fatigue-related declines in luggage-screening performance. *Journal of Occupational and Environmental Medicine*, *53*(10), 1146-1154.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*, 289–300.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, *1*(8476), 307-310.

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8*(2), 1350-1160.

Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: A New Instrument for Psychiatric Practice and Research. *Journal of Psychiatric Research*, *28*(2), 193–213.

Chuah, Y. M. L., Venkatraman, V., Dinges, D. F., & Chee, M. W. L. (2006). The neural basis of interindividual variability in inhibitory efficiency after sleep deprivation. *Journal of Neuroscience*, *26*(27), 7156-7162. doi: 10.1523/JNEUROSCI.0906-06.2006

Combs, E. (2018). *Comparison of physiological and cognitive performance in F-22 pilots during the transition from day to night flying operations*. Naval Postgraduate School, Monterey, CA.

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

Dinges, D. F., & Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments, & Computers*, *17*(6), 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*(3), 253–267.

Drummond, S. P. A., Paulus, M. P., & Tapert, S. F. (2006). Effects of two nights sleep deprivation and two nights recovery sleep on response inhibition. *Journal of Sleep Research*, *15*(3), 261-265.

Durmer, J. S., & Dinges, D. F. (2005). Neurocognitive Consequences of Sleep Deprivation. *Seminars in Neurology*, *25*(1), 117–129.

Grant, D. A., Honn, K. A., Layton, M. E., Riedy, S., & 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. *Behavior Research Methods*, *49*(3), 1020-1029.

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields II: Simulation studies. *Psychophysiology, 48*(12), 1726-1737.

Honn, K. A., Riedy, S., & Grant, D. A. (2015). Validation of a portable, touch-screen psychomotor vigilance test. *Aerospace Medicine and Human Performance, 86*(5), 428-434.

Jewett, M. E., Dijk, D. J., Kronauer, R. E., & Dinges, D. F. (1999). Dose-response relationship between sleep duration and human psychomotor vigilance and subjective alertness. *Sleep*, *22*(2), 171–179.

Kay, M., Grandner, M. A., Bauer, J., Lang, R. A., Watson, N. F., & Kientz, J. A. (2013). Initial validation of an Android-based psychomotor vigilance task. *Sleep*, *36*(Abstract Supplement), A108.

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*. Paper presented at the 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops, Venice, Italy.

Lamond, N., Dawson, D., & Roach, G. (2005). Fatigue Assessment in the Field: Validation of a Hand-Held Electronic Psychomotor Vigilance Task. *Aviaton Space Environmental Medicine, 76*(5), 486-489.

Matsangas, P., & McCauley, M. E. (2014). Sopite syndrome: A revised definition. *Aviation Space and Environmental Medicine*, *85*, 672-673.

Matsangas, P., McCauley, M. E., & Becker, W. (2014). The effect of mild motion sickness and sopite syndrome in cognitive multitasking performance. *Human Factors*, *56*, 1124-1135.

Matsangas, P., & Shattuck, N. L. (2018). Agreement between the 3-minute Psychomotor Vigilance Task (PVT) embedded in a wrist-worn device and the laptop-based PVT. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 666-670.

Matsangas, P., & Shattuck, N. L. (2020). Hand-held and wrist-worn field-based PVT devices vs. the standardized laptop PVT. *Aerospace Medicine and Human Performance*, *91*(5), 409–415. doi: 10.3357/AMHP.5567.2020

Matsangas, P., Shattuck, N. L., & Brown, S. (2017). Preliminary validation study of the 3-minute wrist-worn Psychomotor Vigilance Test. *Behavior Research Methods, 49*(5), 1792-1801. doi: 10.3758/s13428-016-0821-2

Matsangas, P., Shattuck, N. L., Mortimore, K., Paghasian, C., & Greene, F. (2019). The 3-minute Psychomotor Vigilance Task (PVT) Embedded in a Wrist-worn Device: Time of Day Effects. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 63*(1), 797-801. doi: 10.1177/1071181319631144

Mullaney, D. J., Kripke, D. F., & Messin, S. (1980). Wrist-Actigraphic Estimation Of Sleep Time. *Sleep, 3*(1), 83-92.

Ocano, D., Watson, N. F., Kay, M., Kientz, J. A., & Grandner, M. A. (2017). Validation of a touchscreen psychomotor vigilance task for Android devices. *Sleep, 40*(Abstract Supplement), A88.

Sagaspe, P., Taillard, J., Amiéva, H., Beck, A., Rascol, O., Dartigues, J.-F., . . . Philip, P. (2012). Influence of age, circadian and homeostatic processes on inhibitory motor control: A Go/No-Go task study. *PLoS ONE, 7*(6), e39410. doi: 10.1371/journal.pone.0039410

Shattuck, N. L., Matsangas, P., & Brown, S. (2015). *A comparison between the 3/9 and the 5/10 watchbills* (Report No. NPS-OR-15-006). Monterey, CA: Naval Postgraduate School.

Shattuck, N. L., Matsangas, P., & Powley, E. H. (2015). *Sleep patterns, mood, psychomotor vigilance performance, and command resilience of watchstanders on the "five and dime" watchbill* (Technical Report No. NPS-OR-15-003). Monterey, CA: Naval Postgraduate School.

Shattuck, N. L., Matsangas, P., & Waggoner, L. (2014, October 27-31). *Assessment of a novel watchstanding schedule on an operational US Navy vessel.* Paper presented at the Human Factors and Ergonomics Society (HFES) 58th Annual Meeting, Chicago, IL.

Shattuck, N. L., Waggoner, L. B., Young, R. L., Smith, C. S., & Matsangas, P. (2014). Shiftwork practices in the United States Navy: A study of sleep and performance in watchstanders aboard the USS Jason Dunham. *Sleep, 37*(Abstract Supplement), A78.

Thorne, D., 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*(1), 111-118.

Van Wert, M. J., Horowitz, T. S., & Wolfe, J. M. (2009). Even in correctable search, some types of rare targets are frequently missed. *Attention, Perception, and Psychophysics*, *71*(3), 541-553.

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Cognitive Psychology: Rare items often missed in visual searches. *Nature*, *435*(7041), 439-440.

Wyatt, J. K., Dijk, D. J., Ronda, J. M., Jewett, M. E., Powell, J. W., Dinges, D. F., & Czeisler, C. A. (1997). Interaction of circadian- and sleep/wake homeostatic-processes modulate psychomotor vigilance test (PVT) performance (Abstract). *Sleep Research*, *26*, 759.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Research Sponsored Programs Office, Code 41
   Naval Postgraduate School
   Monterey, California

4. Nita Lewis Shattuck………………………………………………………………...1
   Operations Research Department
   Naval Postgraduate School
   Monterey, California

5. Panagiotis Matsangas……………………………………………………………….1
   Operations Research Department
   Naval Postgraduate School
   Monterey, California

THIS PAGE INTENTIONALLY LEFT BLANK