Hand-Held and Wrist-Worn Field-Based PVT Devices vs. the Standardized Laptop PVT

Panagiotis Matsangas; Nita Lewis Shattuck

BACKGROUND: Given the challenges of collecting reliable Psychomotor Vigilance Task (PVT) data in the field, this study compared a 3-min PVT on a hand-held device and wrist-worn device vs. a standardized laptop.

METHODS: The experiment utilized a randomized, repeated-measures design. Subjects ($N = 36$) performed the PVT on a touch-screen, hand-held device (HHD), a wrist-worn device (WWD), and a standardized laptop (L). Sleep was assessed using wrist-worn actigraphy.

RESULTS: Compared to the L, the HHD was slower on average (~50% longer reaction times; ~34% slower response speeds; ~600% more lapses in attention combined with false starts) and introduced a proportional bias that decreased the range of response speeds by 60%. Compared to the L, the WWD with the backlight on was faster on average (reaction time: ~6%; response speed: ~13%), but equivalent in lapses combined with false starts, and introduced a proportional bias that increased the range of responses by 60%.

DISCUSSION: Compared to the L PVT, using a hand-held, touch screen interface to collect PVT data may introduce a large constant bias and a proportional bias that decreases the range of response speed. However, performance on the WWD closely mirrors performance on the L PVT and the proportional bias tends to be in favor of detecting individuals with slower responses. Researchers should avoid comparing PVT metrics between different device types. Reliability of PVT data from a WWD or HHD may be degraded when used in an operational setting with unpredictable environmental movement (such as a surface maritime setting).

KEYWORDS: touch screen devices, psychomotor vigilance performance, Bland-Altman method.

Matsangas P, Shattuck NL. *Hand-held and wrist-worn field-based PVT devices vs. the standardized laptop PVT*. Aerosp Med Hum Perform. 2020; 91(5):409–415.

The PVT is considered the de facto gold standard measure of neurobehavioral effects of sleep loss and circadian misalignment.[5] This task is typically administered on a computer or laptop in controlled laboratory settings. However, application of this task in field-based research is more challenging. In research conducted aboard United States Navy ships, crewmembers were required to perform the PVT prior to and following their shifts. Subjects were required to go to an assigned area aboard the ship and perform the PVT on specifically designated laptops four times per day. Unpublished observations found crewmember compliance to be less than 10%. Time limitations due to additional duties and personal fatigue were identified as key contributing factors to these poor compliance rates. A wearable or hand-held device would prove very beneficial in field-based testing environments.

Previous research investigating the validity of a 3-min PVT, with interstimulus intervals (ISI) of 2–10 s, embedded in a wrist-worn device, found that PVT data could be reliably collected in a field-based environment.[14,15] Results from multiple studies conducted by the NPS Crew Endurance Team aboard ships have shown the utility of collecting PVT data from wrist-worn devices.[19,20] Specifically, the use of actiwatches to collect PVT data increased compliance with the research protocol to approximately 60% (i.e., ~sixfold improvement compared to the use of PVT laptops aboard ships). Furthermore, PVT data collected with the actiwatches in naval settings permitted the

From the Operations Research Department, Naval Postgraduate School, Monterey, CA, USA.

This manuscript was received for review in December 2019. It was accepted for publication in February 2020.

Address correspondence to: Panagiotis Matsangas, Ph.D., 1411 Cunningham Drive, Operations Research Department, Naval Postgraduate School, Monterey, CA 93943; pmatsang@nps.edu.

DOI: https://doi.org/10.3357/AMHP.5567.2020

researchers to objectively distinguish circadian-aligned from noncircadian watchstanding schedules used in the United States Navy. It also aided researchers in identifying more and less fatigued sailors as well as identify differences in reaction time and performance at different times of day.[14,16]

Given the widespread use of hand held devices (HHD) with a touch screen interface, a number of research efforts have explored the utility of a HHD to collect PVT data. Honn and colleagues investigated the utility of a 5-min PVT (ISI = 2–10 s) (Boeing Company, Seattle, WA) installed in a touch screen Personal Digital Assistant (PDA – 3rd generation Apple iPod).[10] Results showed that change in the reaction time (RT) distribution due to sleep deprivation was more substantial on the laptop than on the PDA (134 ms vs. 76 ms). Furthermore, change in spread of the RT distribution due to sleep deprivation was more substantial on the laptop than on the PDA (466 ms vs. 238 ms). Unprecedented false starts were also evident in the PDA among alert individuals. The authors attributed this difference due to the ease of accidental touch on a touch screen device compared to the effortful action needed to press down on a keyboard.[10] Kay and colleagues developed a touch screen version of the PVT ("PVT Touch") implemented in an Android-based device (5-min and 10-min durations; ISI = 2–10 s) and compared it to the validated 10-min PVT-192.[11,12] Results showed that the mean and median RT were higher and lapses in attention more frequent in the touch screen device compared to the PVT-192.[18] Additionally, a 5-min NASA-PVT (ISI = 2–10 s) was developed for touch-screen devices and was implemented in a 5th generation Apple iPod.[2,3] Compared to the original PVT-192, the mean RT was higher and lapses in attention were more frequent in the touch screen device. False starts were not assessed in this study. Another study assessed the utility of a 3-min PVT (Pulsar Informatics, Inc., ISI = 2–5 s) in a smartphone (Galaxy S3 model GT-I9300) and in a tablet (iPad mode A1432, Apple Inc., Cupertino, CA) compared to the 10-min laptop-based PVT (ISI = 2–10 s).[9] In contrast to other studies, results showed that the mean RTs were faster on the smartphone and tablet than on the laptop. False starts in the smartphone were comparable to the laptop, whereas the tablet had fewer false starts. The authors noted, however, that all devices were "individually calibrated" without providing more information regarding this calibration.

Overall, the aforementioned studies indicate that performance on the PVT embedded in touch screen devices (either in the form of a 3-min duration task with ISI = 2–5 s or 5-min duration task with ISI = 2–10 s) is sensitive to wakefulness. While several limitations exist with use of touch screen devices as a method of measuring PVT performance (e.g., larger reaction times, more false starts, and more lapses in attention) compared to the standardized laptop PVT, it was concluded that the use of a PVT + touch screen device may be a valid instrument for measuring the effects of fatigue.

The aim of this study is to assess the utility of a 3-min PVT (ISI = 2–5 s) embedded in devices appropriate for field research (i.e., a touch-screen, hand-held device and a wrist-worn device) using the Bland-Altman method. The primary goal was to compare the PVT embedded in a hand-held device (HHD) with a touch-screen (tablet) with the validated laptop PVT. The secondary goal was to contrast PVT metrics between the HHD and the WWD.

METHODS

Subjects

There were 36 healthy individuals (aged 19–47 yr) from the Naval Postgraduate School (NPS) who volunteered to participate in this study. Informed consent was obtained from each subject prior to any data collection. Ethical approval was granted for this study by the NPS Institutional Review Board. Subjects were screened for corrected vision, recent injuries or pain in the arms, wrists, or fingers, a diagnosis of color vision deficiency, and carpal tunnel syndrome. Subjects slept on average 6.80 ± 0.75 h/d. All subjects had normal daytime sleepiness before both data collection sessions as assessed by the Epworth Sleepiness Scale.

Equipment

A study questionnaire was developed that included demographic questions, sleep history for the 48-h prior to data collection, current day's caffeine intake, and factors that could affect participation in the study. Psychomotor vigilance performance data was collected on three devices: 1) the validated PVT (implementation version 2.0.5.9–Pulsar Informatics Inc., Philadelphia, PA) installed on two individually-calibrated Latitude E6420 laptops with 14″ displays (Dell Inc., Round Rock, TX) running a Windows 7 operating system (Microsoft Inc., Redmond, WA); 2) the PVT-Touch[11,12] installed on two Samsung Galaxy Note 8.0 GT-N5110 (Samsung Electronics Co., Ltd., Suwon, South Korea) tablets serving as the touch-screen HHDs; and 3) the WWD (Motionlogger Watch) with an embedded version of the PVT (Ambulatory Monitoring, Inc., Ardsley, NY). Applications and Wi-Fi were turned off to minimize variability. In all devices, the duration of the PVT trials was 3 min with ISI = 2–5 s. Based on findings from previous research, subjects were instructed to use the touch down technique to respond to the PVT stimulus.[12] In contrast to other PVT studies,[3,9] subjects were instructed to perform the HHD PVT using their dominant hand while holding the device on a table in a portrait orientation.

Sleep was assessed by wrist-worn actigraphy (i.e., AMI Motionlogger which was also used for the PVT) assisted by paper-based activity logs. Data were collected in 1-min 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 5 min. The sleep latency criterion was no more than 1 min awake in a 20-min period (all values were defaults for this software).

Procedures

This study employed a randomized, within-subject, repeated-measures design. Data were collected during two consecutive days in a normal office lighting environment (300–400 lux). Subjects arrived to the laboratory 2 d prior to the first day of data collection. They were issued an actiwatch and an activity log to assess their sleep patterns before and during the study (~3 d in total).

After completing the study questionnaire, subjects were shown how to perform the PVT. Subjects were instructed to respond as soon as each stimulus appeared, but not to anticipate the target because that would yield a false start. Subjects were then permitted one test trial with each device.

The data collection was divided between two consecutive days to avoid boredom and lack of focus during testing. On Day 1, subjects performed three 3-min PVT trials, one on a laptop (L), one on the WWD with the screen red backlight off ($WWD_{BL = OFF}$), and one on the WWD with the screen red backlight on ($WWD_{BL = ON}$). Subjects were randomly assigned to one of the six treatment groups (**Fig. 1**). There was a 1-min break between trials. While performing the tests, subjects were seated and wearing headphones to attenuate ambient noise. A researcher was present behind the subject in the experimentation room to monitor the study. On Day 2, subjects reported to the laboratory at approximately the same time as Day 1 of the experiment and were randomly assigned to a treatment group. Subjects completed the ESS, and performed two 3-min PVT trials, one on the HHD and one on the L (Fig. 1).

Statistical Analysis

A PVT response was regarded as valid if the reaction time (RT) was greater than or equal to 100 ms and less than 30 s. Responses with RTs less than 100 ms were identified as false starts (errors of commission). Lapses in attention were defined as RTs greater than or equal to 355 ms and 500 ms. Based on the PVT metrics proposed by Basner and Dinges,[4] analysis included these PVT metrics: mean reaction time (RT); mean response speed (i.e., reciprocal reaction time, calculated as $1/RT \cdot 1000$ and measured in $10^3 \cdot ms^{-1}$); fastest 10% RT; slowest 10% 1/RT; percentage of false starts; percentage of 355-ms lapses in attention; percentage of 500 ms lapses in attention; percentage of 355-ms lapses in attention combined with false starts; percentage of 500-ms lapses in attention combined with false starts. For all metrics, the response values were aggregated by trial.

First, the data were assessed for normality using the Shapiro-Wilk W test. With the exception of response speed, the PVT data were not normally distributed. Following this assessment, the average difference between devices was assessed. That is, the WWD with the backlight feature off/on was compared with the L PVT on Day 1, and the HHD was compared with the L PVT on Day 2. Multiple comparisons were based on the Tukey-Kramer 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.

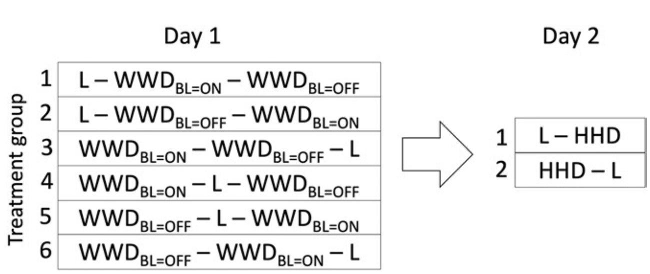

Fig. 1. Experiment design and treatment groups.

The Bland–Altman method was used to assess the agreement between the WWD and HHD PVT systems with the validated L PVT.[1,7,8] 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,[1] The regression approach for nonuniform differences was used if the mean difference between devices was associated with the magnitude of the measurements,[8]

Statistical analysis was conducted with JMP statistical software (JMP Pro 14.2; SAS Institute; Cary, NC). Normally distributed data are presented as M ± SD, whereas nondistributed data are presented as median ± interquartile range (MD ± IQR). An alpha level of 0.05 was used to determine statistical significance. Post hoc statistical significance was assessed using the Benjamini–Hochberg False Discovery Rate (BH-FDR) controlling procedure with $q = 0.20$.[6]

RESULTS

Fig. 2 shows the scatter plots of the response speed of the PVT performed on the L, the WWD with the screen backlight feature on and off, and the HHD. The equivalence line is based on L data. Visual inspection of the WWD data shows that when the backlight is on, response speed is faster (higher) on the WWD compared to the L. When the backlight is off, however, WWD response speed is slower compared to the L. The latter pattern is also evident in the HHD. In terms of dispersion, response speed had a range of ~1.8 $10^3 \cdot ms^{-1}$ in the L and ~2.9 $10^3 \cdot ms^{-1}$ in the WWD, but only ~0.7 $10^3 \cdot ms^{-1}$ in the HHD. That is, the WWD transformed response speed such that the range of values increased by 60% compared to the L. In contrast, the range of responses speed in the HHD decreased by 60% compared to the L.

Initially, the magnitude of the disagreement between devices, expressed as the average differences, was assessed. Compared to the L PVT, the WWD PVT when the backlight is off was associated with higher reaction times, slower response speeds, and more lapses in attention/lapses in attention combined with false starts. In contrast, the WWD with the backlight on was faster than the L PVT in terms of reaction time (Δ=~6%) and response speed (Δ=~13%), but equivalent to the L PVT in terms of lapses in attention combined with false starts. Compared to the L, the PVT embedded in the HHD differed in all PVT metrics, i.e., the HHD PVT was slower with more lapses in attention/lapses in attention combined with false starts. Detailed results for all PVT metrics are presented in **Table I**. Of note, the standard deviation of HHD response speed was approximately half that of the L PVT.

Fig. 3 shows the Bland-Altman plots for the absolute (diagrams A, B, C) and percentage-wise (diagrams D, E, F)

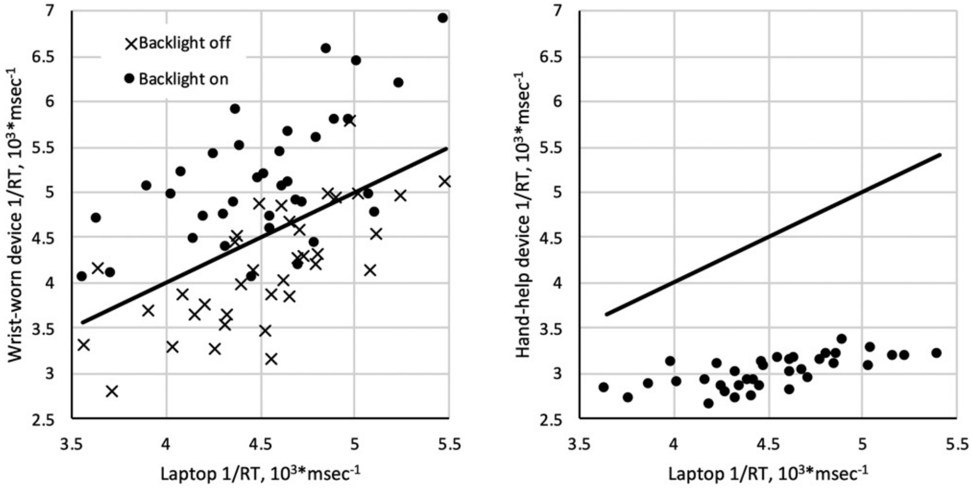

Fig. 2. Wrist-worn device (WWD) and hand-held device (HHD) PVT response speed (1/RT) compared to laptop (L) PVT. The indicated line of equivalence is based on the data from the laptop (L) PVT.

differences between devices. 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 RS were found to be in opposing directions when the backlight feature in the WWD was on. That is, fast individuals tended to perform better in the WWD than the L, whereas slow individuals tended to perform worse in 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 Fig. 3 in conjunction with the HHD scatterplot in Fig. 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).

DISCUSSION

The first aim of this research was to assess the agreement of the 3-min PVT (ISI = 2–5 s) embedded in a touch-screen HHD and the validated L PVT. This study's findings agree with earlier research regarding the large average differences between the HHD and the L PVT metrics.[18] In the present study, these differences represent ~50% larger reaction times and ~34% slower response speeds. The large reaction times in the HHD led to a large number of 355-ms lapses in attention (~29 lapses in attention in the ~48 responses in a 3-min PVT trial). Consequently, 355-ms lapses in attention combined with false starts were higher by ~600% (or ~210% when lapses in attention are calculated by the 500-ms criterion). These findings are not unexpected if one considers that the median reaction time in the HHD was 346 ms which is close to the 355-ms lapse in attention criterion. The fact that, even though subjects were rested, approximately 60% of the responses in the HHD were classified as "lapses in attention," raising a concern regarding the usefulness of the 355-ms lapse criterion for detecting fatigued individuals in field settings. This finding leads to questions regarding the sensitivity of the HHD PVT system to detect fatigued individuals in "fatigue-saturated" field-based environments.

Table I. Comparisons of PVT Metrics Between the Wrist-Worn Device and the Day 1 Laptop PVT, and Between the Hand-Held Device and the Day 2 Laptop PVT.

	DAY 1			DAY 2	
PVT METRIC	**LAPTOP**	**WWD (BACKLIGHT OFF)**	**WWD (BACKLIGHT ON)**	**LAPTOP**	**HHD**
RT (ms), MD ± IQR‡	226 ± 29.0	256 ± 73.5^{A***}	213 ± 41.5$^{B***, C**}$	229 ± 31.1	346 ± 45.7^{D***}
1/RT (1000/ms), M ± SD†	4.52 ± 0.44	4.16 ± 0.66^{A***}	5.11 ± 0.71$^{B***, C***}$	4.52 ± 0.41	2.99 ± 0.18^{D***}
Fastest 10% RT (ms), MD ± IQR‡	191 ± 21.3	199 ± 40.8^{A*}	157 ± 21.4$^{B***, C***}$	191 ± 19.7	287 ± 26.1^{D***}
Slowest 10% 1/RT (1000/ms), MD ± IQR‡	3.65 ± 0.58	3.03 ± 1.10^{A***}	3.46 ± 1.14^{C***}	3.72 ± 0.53	2.54 ± 0.35^{D***}
False Starts (FS) (%), MD ± IQR‡	2.08 ± 4.08	0 ± 2.21	0 ± 2.07^{B**}	2.08 ± 3.62	3.22 ± 4.39^{D*}
Lapses 355ms, (%), MD ± IQR‡	0 ± 2.12	7.69 ± 10.4^{A***}	2.28 ± 6.32$^{B*, C**}$	0 ± 2.08	21.1 ± 22.3^{D***}
Lapses 500ms, (%), MD ± IQR‡	0 ± 0	2.22 ± 4.52^{A***}	0 ± 2.16$^{B**, C*}$	0 ± 4.17	2.15 ± 6.22^{D***}
Lapses 355ms + FS, (%), MD ± IQR‡	2.13 ± 3.55	8.70 ± 11.1^{A***}	4.13 ± 5.70^{C**}	4.13 ± 4.14	28.9 ± 23.8^{D***}
Lapses 500ms + FS, (%), MD ± IQR‡	2.08 ± 3.58	4.09 ± 4.24^{A**}	2.09 ± 4.08^{C**}	2.13 ± 3.62	6.60 ± 7.70^{D***}

M ± SD: Mean ± SD; MD ± IQR: Median ± Interquartile range.
† Pairwise comparisons based on the 2-sided *t*-test for matched pairs; ‡Pairwise comparisons based on the 2-sided Wilcoxon signed rank test.
^{A} 1st day difference between "WWD with backlight off" and "Laptop."
^{B} 1st day difference between "WWD with backlight on" and "Laptop."
^{C} 1st day difference between "WWD with backlight off" and "WWD with backlight on."
^{D} 2nd day difference between "HHD" and "Laptop."
Statistical significance for differences: $*P < 0.05$; $**P < 0.01$; $***P < 0.001$.
Post hoc statistical significance assessed with the Benjamini–Hochberg False Discovery Rate BH-FDR controlling procedure.

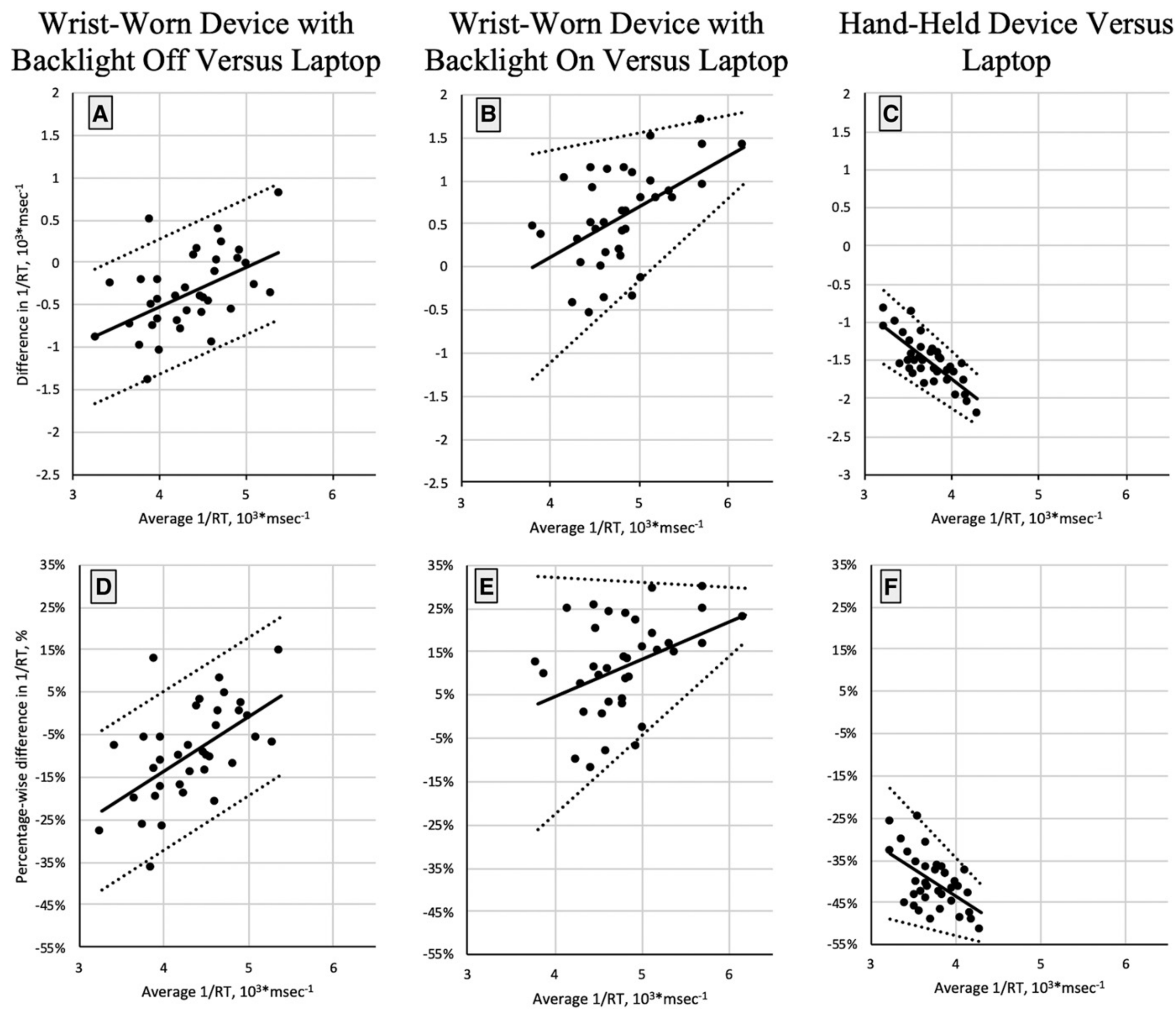

Fig. 3. Bland-Altman plots for response speed. The thick continuous line represents the regression line of the absolute and percentage-wise differences. The 95% limits of agreement are represented by the dotted lines.

To overcome these issues, some researchers have proposed adjusting the HHD PVT results by the average latency of the HHD device.[3] The validity of this approach, however, is based on two assumptions. The first assumption proposes that the device latency (i.e., the time between tapping the screen and the system's response) is constant and known a priori for the specific device. This assumption means that researchers are using specific devices to collect PVT data which have been assigned to this purpose (i.e., the use of software applications other than the PVT are not used – because even updates to the operating system may affect device latency). The second assumption proposes that performing the PVT on the HHD does not introduce any other systematic transformation of the PVT responses. However, the present research demonstrates that the HHD introduces a proportional bias that decreases the range of response speeds in the HHD by 60% compared to the validated L PVT. That is, HHD responses seem to be more clustered than in the L PVT.

In contrast to the issues identified for the HHD, the WWD appeared to have better agreement with the validated L PVT in terms of the magnitude of average disagreement. Specifically, response speed in the WWD with the backlight on was approximately 13% faster compared to the L PVT (the difference was ~6% for reaction time), but the two devices were equivalent in terms of lapses in attention combined with false starts. Even the proportional bias introduced to the PVT responses by the WWD tended to favor detecting individuals with slower responses. That is, the WWD increased the range of responses by 60% compared to the L because the differences in response speed between the WWD and the L were incrementally associated with the magnitude of the response speed. Therefore, response speed in the WWD tended to be faster compared to the L for faster individuals, whereas slower individuals tended to do worse in the WWD compared to the L. These findings are in agreement with previous research in which the WWD with the 3-min PVT with ISI between 2 and 10 s was utilized.[14]

Of note, one issue of concern was that while not clearly noted in the PVT literature, the adjustment of PVT metrics by a fixed value is needed if PVT performance is to be compared with data obtained from different devices. In the case of devices with proportional bias, however, the adjustment approach becomes more complex because of the rigorous and time-consuming testing for each device type to be used in a study. Hence, this approach may not be practical in reality. In such conditions, a more realistic approach would be not to compare PVT performance with data collected with different devices.

Another issue of concern is the interface that a field PVT system should have to collect reliable PVT data in a moving environment. In the present study, subjects hovered their finger a small distance from the screen while waiting for the stimulus. For example, if PVT data were collected aboard a ship while underway, ship motion could lead to unwanted responses and result in motion induced interruptions to task performance.[17] Given that most researchers emphasize the utility of touch screen devices for collecting PVT data, future efforts should assess the effect of environmental motion to PVT metrics when collecting data in the field.

In conclusion, using a hand-held device with a touch screen interface to collect PVT data may introduce a large constant bias and a proportional bias that decreases the range of response speed compared to the validated L PVT. The WWD, however, seems to have a better agreement with the validated L PVT. Even the proportional bias of the WWD tends to favor detection of individuals with slower responses. The following findings should be taken in to consideration when interpreting PVT results; comparing PVT data between different device types should be avoided. Lastly, appropriate methods should be used to assess agreement between measurement devices.

Several limitations should be considered when interpreting this study's findings. This study assessed the differences in PVT performance between devices in a relatively young and healthy population (75% were younger than 34 yr). An additional calibration device was not employed to assess the accuracy of the timed reactions.[13] Different devices with different screen characteristics and stimulus presentation may yield different results.

A short interstimulus interval (ISI) of 2 to 5 s was employed in this research. This ISI leads to a high presentation rate of stimuli, one of the criticisms of the PVT. In general, vigilance tasks in operational environments are characterized by infrequent, or even rare, occurrences that would be better represented by longer ISIs.[21,22] In contrast, much of the PVT literature focuses on ISIs of 1 to 10 s in length. Future studies should assess the utility of PVT with longer ISIs. Lastly, the repeatability of the wrist-worn and the hand-held devices (i.e., variation in repeated measurements on the same subject under the same conditions) was not determined.

ACKNOWLEDGMENTS

This study was supported by funding from the United States Navy Naval Medical Research Center's Advanced Medical Development Program. The authors thank Anna Donnla O'Hagan, Ph.D. for proofreading the manuscript.

Financial Disclosure Statement: The authors have no competing interests to disclose.

Authors and affiliations: Panagiotis Matsangas, Ph.D., and Nita Lewis Shattuck, Ph.D., Research Department, Naval Postgraduate School, Monterey, CA, USA.

REFERENCES

1. Altman DG, Bland JM. Measurement in medicine: The analysis of method comparison studies. Statistician. 1983; 32(3):307–317.
2. Arsintescu L, Kato KH, Cravalho PF, Feick NH, Stone LS, Flynn-Evans EE. Validation of a touchscreen psychomotor vigilance task. Accid Anal Prev. 2019; 126:173–176.
3. Arsintescu L, Mulligan JB, Flynn-Evans EE. Evaluation of a psychomotor vigilance task for touch screen devices. Hum Factors. 2017; 59(4):661–670.
4. Basner M, Dinges DF. Maximizing sensitivity of the Psychomotor Vigilance Test (PVT) to sleep loss. Sleep (Basel). 2011; 34(5):581–591.
5. Basner M, Hermosillo E, Nasrini J, McGuire S, Saxena S, et al. Repeated administration effects on Psychomotor Vigilance Test (PVT) performance. Sleep. 2018; 41(1). doi: 10.1093/sleep/zsx187.
6. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol. 1995; 57:289–300.
7. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986; 327(8476): 307–310.
8. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999; 8(2):135–160.
9. Grant DA, Honn KA, Layton ME, Riedy S, Van Dongen HPA. 3-minute smartphone-based and tablet-based psychomotor vigilance tests for the assessment of reduced alertness due to sleep deprivation. Behav Res Methods. 2017; 49(3):1020–1029.
10. Honn KA, Riedy S, Grant DA. Validation of a portable, touch-screen psychomotor vigilance test. Aerosp Med Hum Perform. 2015; 86(5):428–434.
11. Kay M, Grandner MA, Bauer J, Lang RA, Watson NF, Kientz JA. Initial validation of an Android-based psychomotor vigilance task. Sleep. 2013; 36(Abstract Supplement):A108.
12. Kay M, Rector K, Consolvo S, Greenstein B, Wobbrock JO, et al. PVT - Touch: Adapting a reaction time test for touchscreen devices. 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops. 5 May – 5 Aug 2013; Venice, Italy. 2013:248–251.
13. Khitrov MY, Laxminarayan S, Thorsley D, Ramakrishnan S, Rajaraman S, et al. PC-PVT: A platform for psychomotor vigilance task testing, analysis, and prediction. Behav Res Methods. 2014; 46(1):140–147.
14. Matsangas P, Shattuck NL. Agreement between the 3-minute Psychomotor Vigilance Task (PVT) embedded in a wrist-worn device and the laptop-based PVT. Proc Hum Factors Ergon Soc Annu Meet. 2018; 62(1):666–670.
15. Matsangas P, Shattuck NL, Brown S. Preliminary validation study of the 3-minute wrist-worn Psychomotor Vigilance Test. Behav Res Methods. 2017; 49(5):1792–1801.
16. Matsangas P, Shattuck NL, Mortimore K, Paghasian C, Greene F. The 3-minute Psychomotor Vigilance Task (PVT) Embedded in a Wrist-worn Device: Time of Day Effects. Proc Hum Factors Ergon Soc Annu Meet. 2019; 63(1):797–801.
17. McCauley ME, Matsangas P, Miller NL. Motion and fatigue study in high speed vessel operations: Phase 1 report. Monterey, CA: Naval Postgraduate School; 2005.

18. Ocano D, Watson NF, Kay M, Kientz JA, Grandner MA. Validation of a touchscreen psychomotor vigilance task for Android devices. Sleep. 2017; 40(Abstract Supplement):A88.
19. Shattuck NL, Matsangas P. Operational assessment of the 5-h on/10-h off watchstanding schedule on a US Navy ship: Sleep patterns, mood, and psychomotor vigilance performance of crew members in the nuclear reactor department. Ergonomics. 2016; 59(5):657–664.
20. Shattuck NL, Matsangas P, Brown S. A comparison between the 3/9 and the 5/10 watchbills. Monterey, CA: Naval Postgraduate School; 2015 February. Report No.: NPS-OR-15-006.
21. Van Wert MJ, Horowitz TS, Wolfe JM. Even in correctable search, some types of rare targets are frequently missed. Atten Percept Psychophys. 2009; 71(3):541–553.
22. Wolfe JM, Horowitz TS, Kenner NM. Cognitive Psychology: Rare items often missed in visual searches. Nature. 2005; 435(7041):439–440.