MOUNTAIN-PLAINS CONSORTIUM

MPC 21-432 | J. Owen and P. Sherry

Validation of AlertMeter®
Fatigue Assessment Device
for Transportation Workers

A University Transportation Center sponsored by the U.S. Department of Transportation serving the Mountain-Plains Region. Consortium members:

Colorado State University
North Dakota State University
South Dakota State University
University of Colorado Denver
University of Denver
University of Utah
Utah State University
University of Wyoming

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
MPC-605		

4. Title and Subtitle	5. Report Date
Validation of AlertMeter® Fatigue Assessment Device for Transportation Workers	March 2021
	6. Performing Organization Code
7. Author(s) P. Sherry, J. Owen	8. Performing Organization Report No. MPC 21-432
9. Performing Organization Name and Address University of Denver, Denver, CO	10. Work Unit No. (TRAIS)
	11. Contract or Grant No.
12. Sponsoring Agency Name and Address Mountain-Plains Consortium, North Dakota State University, PO Box 6050, Fargo, ND 58108	13. Type of Report and Period Covered: Final Report
	14. Sponsoring Agency Code

15. Supplementary Notes

Supported by a grant from the US DOT, University Transportation Centers Program

16. Abstract

The current study examined a new mobile application for the measurement of fatigue and alertness in the workplace, the AlertMeter®. Seventeen participants experienced 36-hour sleep deprivation and were measured every two hours. Sixteen of the 17 participants were able to stay awake the entire study. No adverse events were observed or reported. The AlertMeter® demonstrated strong validity as a measure of fatigue. It demonstrated significant concurrent validity with the psychomotor vigilance test (PVT) and self-ratings of fatigue (KSS). The AlertMeter® memory tests demonstrated inconsistent correlations with other measures of working memory. Overall, the AlertMeter® is an effective and valid tool for detecting fatigue in the workplace.

17. Key Word	18. Distribution Statement
alertness, commercial drivers, driver performance, evaluation and assessment, fatigue (physiological condition), measurement, smartphones, transit operators, validation	Public distribution

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	50	n/a

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Validation of AlertMeter® Fatigue Assessment Device for Transportation Workers

Jesse Owen, Ph.D.
Professor, University of Denver
Jesse.Owen@du.edu

Patrick Sherry, Ph.D.
Research Professor, NCIT
University of Denver
Patrick.Sherry@du.edu

March 2021

Acknowledgements

The authors would like to thank Marcus Wichman and the staff of Predictive Safety for their sponsorship of this research. In addition, the considerable efforts of Jeremy Coleman, Chrissy Motzny and Sree Sinha of the University of Denver were also greatly appreciated. The support of the Mountain Plains Consortium and Dr. Denver Tolliver is greatly appreciated.

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

TABLE OF CONTENTS

1. INTRODUCTION **1**

2. METHOD **8**

- 2.1 Participants 8
- 2.2 Measures 8
- 2.3 Procedure 9

3. RESULTS **11**

- 3.1 Individual Performance on PVT 16
- 3.2 Main Results 18
  - 3.2.1 Results for AlertMeter®-Mobile, No Memory 19
  - 3.2.2 Results for AlertMeter®-Mobile, with Memory 20
  - 3.2.3 Results for AlertMeter®-Mobile, No Memory by Bins 22
  - 3.2.4 Results for AlertMeter®-Web/Tablet, No Memory 24
  - 3.2.5 Results for AlertMeter®-Mobile, with Memory by Bins 26
  - 3.2.6 Results for AlertMeter®-Web/Tablet, with Memory, by Bins 28
- 3.3 AlertMeter® Memory Correlations with WMTB Tests 30

4. DISCUSSION **36**

- 4.1 AlertMeter® Correlation with the PVT 36
- 4.2 AlertMeter® Stimulus Items 36
- 4.3 Memory – AlertMeter® Memory Correlations with WMBT Tests 37
- 4.4 Study Limitations 38

5. CONCLUSIONS AND RECOMMENDATIONS **39**

6. RECOMMENDATIONS **40**

REFERENCES **41**

LIST OF TABLES

Table 3.1 Descriptive Statistics for AlertMeter® RT, Mobile and Web by Participant 12
Table 3.2 Descriptive Statistics for AlertMeter® RT, Mobile with Memory by Participant ... 14
Table 3.3 Descriptive Statistics for AlertMeter® RT, Web/Tablet with Memory by Participant 15
Table 3.4 AlertMeter® (non-memory) and PVT Growth Models Results 19
Table 3.5 AlertMeter® with Memory and PVT Growth Models Results................................ 21
Table 3.6 AlertMeter® and KSS & PVT Lapses 21
Table 3.7 Correlations of AM Mobile RT (nm) with PVT and KSS by Bin. 23
Table 3.8 Correlations AM Web/Tablet RT (nm) with PVT-RT, PVT-Lapses KSS by Bin 26
Table 3.9 Correlations AM Mobile RT (with Mem) PVT-RT, PVT-Lapses, KSS by Bin ... 28
Table 3.10 Correlations AM Web/Tablet RT (with mem) PVT-RT, PVT-Lapse, KSS by Bin 30
Table 3.11 Average correlations between AlertMeter® and WMTB 31
Table 3.12 Correlations AM Memory and Shape Recognition Accuracy and WMTB Subscales 35

LIST OF FIGURES

Figure 3.1 Means of AM-RT, PVT and KSS by Administration session 13
Figure 3.2 PVT results from Lim & Dinges (2008) 15
Figure 3.3 AlertMeter® Memory AM-RT, PVT-RT, PVT-Lapses, and KSS by Admin sessions 16
Figure 3.4 PVT-RTs for study participants over 36 hours 17
Figure 3.5 AM Web-RT by Individual... 18
Figure 3.6 AlertMeter®-Mobile (non-memory) RT - across tessing sessions......................... 20
Figure 3.7 Means of AM Mobile and Web PVT-RT, PVT-Lapses and KSS 21
Figure 3.8 AM Mobile (nm) RT w Bins 23
Figure 3.9 Mean AM RTs for Web/Table (nm) by Bins, PVT , Lapses and KSS 25
Figure 3.10 Mean RTs of bins for AM-Mobile - Memory and PVT 27
Figure 3.11 Mean RT of significant bins for AM-Tablet - w Memory and PVT 29
Figure 3.12 Memory Accuracy for Mobile Shape Recognition... 32
Figure 3.13 Memory Accuracy for AM Web/Tablet 33
Figure 3.14 Memory Accuracy for AM Mobile.. 33

ABSTRACT

The current study examined a new mobile app for the measurement of fatigue and alertness in the workplace, the AlertMeter®. Seventeen participants experienced 36-hour sleep deprivation and were measured every two hours. Sixteen of the 17 participants were able to stay awake the entire study. No adverse events were observed or reported. The AlertMeter® demonstrated strong validity as a measure of fatigue. It demonstrated significant concurrent validity with the psychomotor vigilance test (PVT) and self-ratings of fatigue (KSS). The AlertMeter® memory tests demonstrated inconsistent correlations with other measures of working memory. Overall, the AlertMeter® is an effective and valid tool for detecting fatigue.

1. INTRODUCTION

Fatigue has been thought to be a major factor contributing to workplace safety, accidents, injuries, and decreased production efficiency. Improving transportation safety is the number one strategic goal of the U.S. Department of Transportation (USDOT) namely to "Reduce transportation-related fatalities and serious injuries across the transportation system."[1] A recent publication from National Highway Traffic Safety Administration (NHTSA) estimates that from 2 percent to 20 percent of annual traffic deaths are attributable to driver drowsiness.[2] According to NHTSA, annually on average from 2009 to 2013, there were over 72,000 police-reported crashes involving drowsy drivers, injuring more than an estimated 41,000 people and killing more than 800, as measured by NHTSA's Fatality Analysis Reporting System (FARS) and National Automotive Sampling System (NASS) General Estimates System (GES).[3] FARS is a census of all fatal crashes that occur on the nation's roadways. NASS GES contains data from a nationally representative sample of police-reported crashes that result in a fatality, injury, or property damage. Using these databases, one study inferred the existence of additional drowsy-driving crashes by looking for correlations with related factors such as the number of passengers in the vehicle, crash time and day of the week, driver sex, and crash type. Similarly, another study was conducted by the AAA Foundation for Traffic Safety and analyzed data from NHTSA's NASS Crashworthiness Data System (CDS). They estimated that 7 percent of all crashes and 16.5 percent of fatal crashes involved a drowsy driver. This suggests that more than 5,000 people died in drowsy-driving-related motor vehicle crashes across the United States in the study year.[4]

More specifically, a fatal bus crash in California's Central Valley in August 2016 was caused by a "severely sleep-deprived driver" and a bus company with an abysmal safety record, according to federal investigators in a report released November 13. Reports describe a severely sleep-deprived driver and a bus company with a poor safety record were causes of the crash that killed four passengers and injured 20 others, including the driver. The NTSB reported the driver had only slept about five hours over the 40 hours preceding the August 2, 2016 crash[5]. The bus, traveling from Los Angeles to Modesto, drifted off the right side of Route 99 and struck a highway signpost that nearly sliced the bus from nose to tail.[2] Fatigue was also cited as a causal factor in a crash that killed 13 people on Interstate 10 near Palm Springs on October 23, 2016, when a charter bus traveling from a casino plowed into the rear of a big-rig whose driver had fallen asleep during a freeway closure. The truck driver was later charged with 13 counts of vehicular manslaughter with gross negligence[3,4]. According to FMCSA records, Autobuses Coordinados vehicles failed eight of 29 federal inspections in just under two years, pushing its out-of-service rate to 38 percent, almost five times greater than the national average of eight percent. These practices, and the presence of fatigued drivers, suggest a failed safety culture that may have contributed to the high levels of fatigued drivers and safety violations.

---

[1] DOT Strategic Plan. https://www.transportation.gov/dot-strategic-plan (from web August 16, 2018.)

[2] 2015 Lifesavers National Conference on Highway Safety Priorities, Mark R. Rosekind, Ph.D., Administrator, NHTSA (March 16, 2015). Retrieved from www.nhtsa.gov/About+NHTSA/Speeches,+Press+Events+&+Testimonies/remarks-mr-lifesavers-03162015

[3] National Center for Statistics and Analysis. (2011, March). Traffic Safety Fact Crash*Stats: Drowsy Driving. (DOT HS 811 449). Washington, DC: NHTSA. Retrieved from www-nrd.nhtsa.dot.gov/pubs/811449.pdf

[4] Tefft, B. C. (2012). Prevalence of motor vehicle crashes involving drowsy drivers, United States, 1999-2008. Accident Analysis & Prevention, 45(1): 180-186.

Similarly, on, January 4, 2017, Long Island Railroad (LIRR) passenger train 2817, consisting of six cars, collided with the platform at the end of Track 6 in the Atlantic Terminal in Brooklyn (a borough of New York City, New York). The lead end of the lead car came to rest on top of the concrete platform at the end of the track. As a result of this accident, 108 people were injured. Damage was estimated at $5.3 million. The NTSB determined the probable cause of the accident was that the engineer fell asleep due to his chronic fatigue as a result of the engineer's severe undiagnosed obstructive sleep apnea.[5]

More recently, Gottlieb, et. al. (2018) studied the relationship between sleep duration and motor vehicle crashes in a sample of 3,201 adults, 222 (6.9%) reported at least one motor vehicle crash during the prior year. Fewer hours of sleep ($p = 0.04$), and self-reported excessive sleepiness ($p < 0.01$) were each significantly associated with crash risk. Severe sleep apnea was associated with a 123% increased crash risk, compared to no sleep apnea. Sleeping six hours per night was associated with a 33% increased crash risk, compared to sleeping seven or eight hours per night. These associations were present even in those who did not report excessive sleepiness. The population-attributable fraction of motor vehicle crashes was 10% due to sleep apnea and 9% due to sleep duration less than seven hours. Thus, poor sleep, due to either sleep apnea or insufficient sleep duration is strongly associated with motor vehicle crashes in the general population, independent of self-reported excessive sleepiness.[6] A meta-analytic review of studies designed to investigate the relationship between sleepiness at the wheel and motor vehicle accidents was conducted in 2017. The authors concluded that drivers experiencing sleepiness at the wheel are at an increased risk of motor vehicle accidents.[7]

In the rail industry, the Hours of Service Law (HSL), first enacted in 1907 and most recently amended in 2008, controls how many hours train, dispatching service, and signal employees may work. The statute provides maximum on-duty periods for each group of employees, minimum off-duty periods for train and signal employees; and establishes how time on duty is to be calculated. The statute also provides additional limitations on consecutive-days and certain monthly limitations on the activity of train employees. In the Rail Safety Improvement Act of 2008, FRA received regulatory authority to establish hours of service limitations for train employees providing commuter and intercity rail passenger transportation service and on August 12, 2011, FRA published its final rule providing new limitations for passenger train employees, which necessitated the evaluation of work schedules for risk of fatigue.

Efforts to identify operators and drivers who might experience fatigue or sleepiness when operating a motor vehicle have been on identifying persons who might experience fatigue in the near future and on monitoring drivers. Technology is under development to assess driver performance, while driving, related to fatigue. However, the need to identify operators fatigue levels and potential for decreased performance due to fatigue *before they begin to operate a vehicle* is also of considerable importance. Once an individual has begun to operate a vehicle

---

[5] NTSB. 2018. https://ntsb.gov/investigations/AccidentReports/Reports/RAB1802.pdf

[6] Gottlieb, D. J., Ellenbogen, J. M., Bianchi, M. T., & Czeisler, C. A. (2018). Sleep deficiency and motor vehicle crash risk in the general population: a prospective cohort study. *BMC medicine*, *16*(1), 44. doi:10.1186/s12916-018-1025-7.

[7] Bioulac, S., Micoulaud-Franchi, J. , Arnaud, M., Sagaspe, P., Moore, N., Salvo, F., Philip, P. (2017). Risk of Motor Vehicle Accidents Related to Sleepiness at the Wheel: A Systematic Review and Meta-Analysis, *Sleep*, Volume 40, Issue 10, October 2017, zsx134, https://doi-org.du.idm.oclc.org/10.1093/sleep/zsx134

even more risk is encountered. Consequently, several attempts to develop devices to measure fatigue outside of vehicles have been attempted.

The psychomotor vigilance test (PVT) (Dinges & Powel, 1985) is perhaps one of the most widely used standard reaction time performance measure. Originally developed for laboratory studies based on a classic psychophysiological reaction time test, it has been used in several studies, including one that observed the sleepiness and fatigue of truck drivers. The PVT can be administered in a standardized format in which the visual stimulus is presented on a computer monitor and the study participant must then respond to having seen the stimulus. The difference between the onset of the presentation of the stimulus and the response is considered the reaction time. The PVT has been shown to be a valid tool for assessing behavioral alertness and vigilant attention performance in a large number of experimental, clinical, and operational paradigms. Balkin et al. [2004] assessed the utility of a variety of instruments for monitoring sleepiness-related performance decrements and concluded that the PVT "was among the most sensitive to sleep restriction, was among the most reliable with no evidence of learning over repeated administrations, and possesses characteristics that make it among the most practical for use in the operational environment."

A review of the existing fatigue detection devices by Dawson (2014) found that none of the current technologies met *all* the proposed regulatory criteria for a legally and scientifically defensible device.[8] Golz, M., Sommer, D., & Trutschel, U. (2010) evaluated commercially available devices for driver fatigue monitoring with particular focus on the needs of the mining industry. Three video-based devices were selected and used with 14 volunteers in an overnight driving simulation study to test their accuracy. EEG and EOG along with percentage of eye closures (PERCLOS), subjectively rated fatigue on the Karolinska Sleepiness Scale (KSS) and driving performance in terms of standard deviation of lateral position in lane (SDL) were also recorded throughout testing sessions. Regression analysis revealed that PERCLOS was significantly related to higher KSS scores and to SDL. The results suggest that under laboratory conditions, current FMT devices are reliable and data averaged across several subjects is utilized but it fails to give a valid prediction of subjective fatigue and of driving performance on an individual level.[9]

In one recent study by Lee et. al (2010), 48 participants completed the polysomnography and the Multidimensional Fatigue Symptom Inventory-short form (MFSI-sf). After sleep monitoring and psychological assessments, the PVT was administered for 10 minutes. Simple correlations and hierarchical linear regression were used to examine the association between PVT lapse count age, apnea hypopnea index (AHI), fatigue, and PVT reaction time. (Lee, Bardwell, Israel, & Dimsdale, 2010). Results showed that PVT lapse count was significantly associated with MFSI-sf physical fatigue ($r = 0.324$, $p = 0.025$). In hierarchical regression, the full model ($R^2 = 0.256$, $p = 0.048$) and higher MFSI-sf physical fatigue ($p = 0.040$) also predicted PVT lapse count. In

---

[8] Dawson, D. (2014) Look before you sleep: Evaluating the use of fatigue detection technologies within a fatigue risk management system for the road transport industry. Sleep Medicine Reviews 18, 2, 2014, 141-152.
[9] GOLZ, M., SOMMER, D., TRUTSCHEL, U. ET AL. EVALUATION OF FATIGUE MONITORING DEVICES. *SOMNOLOGY* (2010) 14: 187. HTTPS://DOI.ORG/10.1007/S11818-010-0482-9

conclusion, the findings suggest that even after controlling for age, BMI, depression, and apnea severity, that fatigue is associated with the PVT lapse.

In another study, eye movements were measured during vigilance tasks following restricted sleep (n = 33 participants) to compare ocular measures to a standard measure of drowsiness (OSLER). Their accuracy was tested for detecting increasing frequencies of behavioral lapses on a different task (psychomotor vigilance task [PVT]). Results indicate that the average duration of eyelid closure and the ratio of the amplitude to velocity of eyelid closure were reliable indicators of frequent errors and detecting ≥ 3 lapses (PVT). The authors concluded that ocular measures, such as duration of episodes of eye closure are promising real-time indicators of drowsiness. (Wilkinson, et. Al., 2013). Wang (2016) also determined that driver drowsiness detection was significantly determined by a combination of percentage of eyelid closure (PERCLOS), average pupil diameter, standard deviation of lateral position and steering wheel reversals.[10]

The need for a more portable phone or smartphone-based app to measure fatigue is apparent. The traditional tried and true laboratory devices are not practical for the workplace or the field. Grant, et. al. (2017) used two versions of the psychomotor vigilance test (PVT) to measure alertness due to sleep loss. Two, 3-minute versions of the psychomotor vigilance test, one smartphone-based and the other tablet based, were validated against a conventional 10-minute laptop-based PVT. Sixteen healthy participants (ages 22–40; seven males, nine females) completed a laboratory study, which included a practice and a baseline day, a 38-hour total sleep deprivation (TSD) period, and a recovery day, during which they performed the three different versions of the PVT every three hours. For each version of the PVT, the number of lapses, mean response time (RT), and number of false starts showed statistically significant changes across the sleep deprivation and recovery days. The number of lapses on the laptop was significantly correlated with the numbers of lapses on the smartphone and tablet. The mean RTs were generally faster on the smartphone and tablet than on the laptop. All three versions of the PVT exhibited a time-on-task effect in RTs, modulated by time awake and time of day. False starts were relatively rare on all three PVTs. For the number of lapses, the effect sizes across 38 hours of TSD were large for the laptop PVT and medium for the smartphone and tablet PVTs. These results indicate that the 3-minute smartphone and tablet PVTs are valid instruments for measuring reduced alertness due to sleep deprivation and restored alertness following recovery sleep.[11]

Brunet (2017) investigated a smartphone app, the Sleep-2-Peak (s2P), that is fairly similar to the PVT for measuring fatigue. To establish validity 3-min versions of s2P and the PVT were administered to participants every two hours during a 35-h total sleep deprivation protocol. The results showed that the s2P successfully distinguished between alert and sleepy states in the same individual and showed decreasing performance as sleep loss increased. Thus, the s2P produces results that can distinguish a sleep-deprived from a non-sleep-deprived individual and is also equivalent to the PVT in detecting performance decrements. Moreover, a strong

[10] Wang,X., & Xu, C. (2016) Driver drowsiness detection based on non-intrusive metrics considering individual specifics. *Accident Analysis & Prevention*, Volume 95, Part B, 2016, pp. 350-357.

[11] Grant, D.A., Honn, K.A., Layton, M.E. et al. Behav Res (2017).

relationship between the s2P and subjective measures of sleepiness, was found similar to findings from other PVT studies (Kaida et al., 2006; Van Dongen et al., 2003).[12]

Price, Moore, Galway, & Linden (2017) attempted to determine whether a smartphone app could be used to assess cognitive functioning in persons who might develop cognitive impairment due to brain injuries. The authors developed a smartphone app to track daily cognitive performance to assess potential levels of cognitive fatigue. Twenty-one participants with no prior reported brain injuries participated in a two-week study. Three cognitive tests were administered six times per day: (1) Spatial Span to measure visuospatial working memory; (2) Psychomotor Vigilance Task (PVT) to measure sustained attention, information processing speed, and reaction time; and (3) a Mental Arithmetic Test to measure cognitive throughput. A smartphone-optimized version of the Mental Fatigue Scale (MFS) self-assessment questionnaire was used as a baseline to assess the validity of the three cognitive tests, as the questionnaire has already been validated in multiple peer-reviewed studies. The PVT showed a positive correlation with the pre-validated MFS $r= 0.342$ ($P<.008$). Scores from the cognitive tests were entered into a regression model and showed that only reaction time in the PVT was a significant predictor of fatigue ($P=.016$, $F=2.682$, 95% CI 9.0-84.2). Higher scores on the MFS were related to increases in reaction time during our mobile variant of the PVT.[13]

Jongstra (2018) also attempted to demonstrate that a smartphone app assessing neuro psychological functioning would be feasible for use with people at increased risk of dementia during a six-month follow-up period. The smartphone-based app iVitality was used to administer neuropsychological tests (Memory-Word, Trail Making, Stroop, Reaction Time, and Letter-N-Back) in healthy adults. Feasibility was tested by studying adherence of all participants to perform smartphone-based cognitive tests. A total of 151 participants (mean age in years=57.3, standard deviation=5.3) used the app with a mean adherence of 60% (SD 24.7) during a 6-month time period. Results showed that there was moderate correlation between the first attempt of the smartphone-based test and the conventional Stroop and Trail Making tests with Spearman $\rho=.3-.5$ ($P<.001$). Correlations increased for both tests when comparing the conventional test with the mean score of all attempts a participant had made, with the highest correlation with Stroop panel 3 ($\rho=.62$, $P<.001$). Performance on the Stroop and the Trail Making tests improved over time suggesting a learning effect, with the other tests remaining stable. Moderate validity for the Stroop and the Trail Making tests compared with conventional neuropsychological tests.[14]

Arsintescu, et. al. (2019) investigated the use of a handheld assessment device to assess fatigue and alertness. Based on the Psychomotor Vigilance Task (PVT), which is considered the gold standard fatigue detection test and is used frequently in fatigue research. Ten participants completed a 5-minute PVT (NASA-PVT) on a touchscreen device and a 5-minute PVT on the

---

[12] Brunet, JF., Dagenais, D., Therrien, M., Gartenberg, D., and Forest, G. (2017) Validation of sleep-2-peak: A smartphone application that can detect fatigue related changes in reaction times during sleep deprivation. Behavioral Research. (2017) 49: 1460. https://doi.org/10.3758/s13428-016-0802-5

[13] Price E, Moore G, Galway L, Linden M. (2017) Validation of a Smartphone-Based Approach to In Situ Cognitive Fatigue Assessment. JMIR Mhealth Uhealth 2017;5(8):e125. URL: https://mhealth.jmir.org/2017/8/e125 DOI: 10.2196/mhealth.6333 PMID: 28818818 PMCID: 5579321.

[14] Jongstra S, Wijsman LW, Cachucho R, Hoevenaar-Blom MP, Mooijaart SP, Richard E Cognitive Testing in People at Increased Risk of Dementia Using a Smartphone App: The iVitality Proof-of-Principle Study. JMIR Mhealth Uhealth 2017;5(5):e68 URL: https://mhealth.jmir.org/2017/5/e68 DOI: 10.2196/mhealth.6939 PMID: 28546139 PMCID: 5465383

original PVT-192. On the day of the experiment, participants arrived in the lab approximately two hours after their habitual wake time. Participants began a routine protocol under dim lighting, beginning two hours after their habitual wake time. The 5-minute PVT-192 and NASA-PVT were taken every two hours for at least 24 hours. The touchscreen NASA-PVT and original computer monitor PVT-192 were sensitive to extended wakefulness in the same manner. The reaction times were slower, and the lapses were higher as time progressed on both NASA-PVT and PVT-192 ($p<0.001$). Also, as expected, reaction time decreased significantly after 16h of wakefulness. Performance continued to deteriorate and was at its worst after 24 hours of wakefulness for both PVTs ($p<0.001$). Thus, the data suggest that the handheld touchscreen device NASA-PVT is a valid tool for assessing fatigue in field studies.[15]

In addition to the effects of fatigue on vigilance and reaction time other studies have demonstrated the relationship between fatigue and memory. Memory function is particularly important in the workplace as orders, instructions, requests, and other procedural knowledge could result in safety and performance decrements if impaired.[16]

In the tasks measuring attention or working memory, two aspects of performance are important: speed and accuracy. In practice, people can switch their emphasis between the two with attentional focusing (Rinkenauer, et. al., 2004). Oftentimes, concentrating on improving one aspect leads to the deterioration of the other. This is called the speed/accuracy trade-off phenomenon. Some SD studies have found impairment only in performance speed, whereas accuracy has remained intact (De Gennaro et al 2001; Chee and Choo 2004). In others, the results are the opposite (Kim, et. al. 2001; Gosselin, et. al. 2005). De Gennaro et. al. (2001) proposed that in self-paced tasks, there is likely to be a stronger negative impact on speed, while accuracy remains intact. In experimenter-paced tasks, the effect would be the opposite. However, many studies show a detrimental effect on both speed and accuracy (e.g., Smith, et. al. 2002; Jennings, et. al. 2003; Chee and Choo 2004; Habeck, et. al. 2004; Choo, et. al. 2005). The speed/accuracy trade-off phenomenon is moderately affected by gender, age, and individual differences in response style (Blatter, et. al. 2006; Karakorpi, et. al. 2006), which could be a reason for inconsistencies in the SD results.[17]

As can be seen from the above review, the need for a quick and portable assessment tool, prior to a person beginning operation of a vehicle would be extremely valuable to the safety of the transportation system. The purpose of the present proposed project is to gather data designed to validate the AlertMeter® as new and promising measure of fatigue and cognitive impairment. The AlertMeter® is a state-of-the-art electronic vigilance test, wherein the stimulus is presented to the study participant visually. The AlertMeter® claims to be able to evaluate the presence of fatigue and cognitive functioning in a very brief period of time. Currently, there are few ultra-brief measures of fatigue/vigilance that have been validated. The AlertMeter® can assess vigilance within two minutes (other assessments can take up to 30 minutes). The proposed

[15] 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 & Prevention*, Volume 126, 2019, Pages 173-176, ISSN 0001-4575, https://doi.org/10.1016/j.aap.2017.11.041.
[16] Alhola, P., & Polo-Kantola, P. (2007). Sleep deprivation: Impact on cognitive performance. *Neuropsychiatric disease and treatment*, *3*(5), 553–567.
[17] Ibid.

project will build on previous research to demonstrate the accuracy of a measure of fatigue and alertness as well as a memory component.

1.1 Research Objectives

The project utilized experimental software from the AlertMeter®, and PVT, Working Memory standardized self-report instruments and Fitbit to assess sleep and wakefulness. The data was analyzed using statistical techniques to review and evaluate the correlation between existing measures (i.e. the PVT) and to demonstrate the occurrence of performance decrements with increase wakefulness. The objectives of this project are as follows:

1) Compare the AlertMeter® to other existing measures of fatigue and alertness.
2) Compare the four versions of AlertMeter® (i.e., smart phone version with memory component, smart phone version without memory component, tablet version with memory component, tablet version without memory component).

2. METHOD

2.1 Participants

We recruited 33 participants from the Denver-Metro area to participate in a fatigue study. Participants were recruited through emails to groups or forums (e.g., Next Door, Craigslist). Initial phone screenings were conducted to ensure that the participants met the inclusion criteria. The most common reason for not participating was time conflicts. After initial screening 17 participants were included in the study. The average age for study participants was 31.18 (*SD* = 8.35; range 24-52). Nine participants identified as female (52.9%) and eight identified as male (47.1%). Participants received: $200 for pre-test assessments, $720 for Sleep Deprivation trial, $200 for post-test assessments, $100 for completion bonus, and $200 for a Fitbit. If they were unable to complete the study, compensation was prorated. Only one participant was not able to stay awake for the full 36-hours.

2.2 Measures

AlertMeter®: The AlertMeter® is a state-of-the-art electronic vigilance test. The stimulus is presented to the study participant visually. The AlertMeter® was administered to the study participants at the beginning of the baseline period commencing three days before the sleep deprivation period to achieve a baseline. During the sleep deprivation part of the study, they were asked to complete the measure on a regular basis (e.g., every two hours). The participants also completed the AlertMeter® in the post phase of the study. The AlertMeter® was administered on two platforms (grid-based platform on a tablet and mobile platform on a smartphone), which were tested with and without a memory module in the test.

Perceptual Vigilance Test (PVT): The PVT (Dinges & Powel, 1985) is perhaps one of the most widely used standard reaction time performance measure. It has been used in several studies including one that looked at the sleepiness and fatigue of truck drivers. The PVT can be administered in a standardized format in which the visual stimulus is presented on a computer monitor and the study participant must then respond to having seen the stimulus. The difference between the onset of the presentation of the stimulus and the response is considered the reaction time. The PVT has been shown to be a valid tool for assessing behavioral alertness and vigilant attention performance in a large number of experimental, clinical, and operational paradigms. Balkin et al. [2004] assessed the utility of a variety of instruments for monitoring sleepiness-related performance decrements and concluded that the PVT "was among the most sensitive to sleep restriction, was among the most reliable with no evidence of learning over repeated administrations, and possesses characteristics that make it among the most practical for use in the operational environment."

Working Memory Test Battery (WMTB): According to Frenda & Fenn (2016) sleep deprivation has been linked to slowed reaction times in simple attention tasks, decreased auditory vigilance and visuospatial attention, and impaired verbal working memory. The bulk of the research suggests that sleep is essential for working memory. Working memory (WM) tasks—and in particular, counting span, operation span, and reading span tasks—are widely used measures of WM capacity. Conway (2004) addressed the reliability and validity of the tasks, and optimal administration and scoring procedures. In the present study, a battery of Working

Memory tasks (WMTB) will be administered to assess changes in working memory following sleep deprivation. A standard set of tasks is suggested by Frenda & Fenn (2016) that can be adapted for this study. Conway (2012) notes "the failure to account for practice effects can have serious implications." Consequently, the battery for the current study will use alternate forms wherever possible. Also, a version of the Digit Span Forward (DSF) and the Digit Span Backwards (DSB) subtests of the WAIS-IV will be modified, according to suggestions by Blackburn & Benton (1957) to improve reliability.

Stanford Sleepiness Questionnaire (SSQ): The SSQ is a standard Likert response scale used to assess the study participant's self-assessment of their sleepiness (Hoddes & Dement, 1972). The SSQ is administered using a nine-point scale. Responses provide an assessment of the level of sleepiness experienced by the participant.

Daily Sleep Log (DSL): A daily sleep log is a self-report instrument constructed to provide information on the level of activity, sleep, and work that a person engages in over the course of a two-week period. The study participant enters data into a self-report booklet designed to monitor activity.

Fitbit: A self-contained device consisting of a computer chip and an accelerometer that provides an estimate of various activities and movements engaged in over the course of a specified time period. The Fitbit device can be calibrated to accept reading in 10, 15, or 30 second time intervals. Results of these calibrations are then set to estimate amount of sleep or activity. We used the Fitbit data to ensure participants were getting 6–8 hours of sleep per night, leading up to the wakefulness phase.

Epworth Sleepiness Scale (ESS): The ESS (Johns, 1991) is a self-administered questionnaire with eight questions. Respondents are asked to rate, on a four-point scale (0-3), their usual chances of dozing off or falling asleep while engaged in eight different activities. The higher the ESS score, the higher that person's average sleep propensity in daily life (ASP), or their "daytime sleepiness." The ES had a statistically significant association with self-rated problem sleepiness. The questionnaire takes no more than two or three minutes to answer.

2.3 Procedure

All research procedures were approved through the University of Denver's IRB (#1304004-1). Data was collected on a sample of healthy adults over a period of 36 hours of wakefulness. Participants were asked to wear a Fitbit and record their sleep patterns using the Denver Sleep Log (DSL) for five days leading up to the wakeful period. In addition, they completed the AlertMeter® and PVT tests multiple times to establish a baseline. At 9 a.m. on Friday morning, the participants checked into the hotel and for the next 36 hours at two hours intervals they were assessed on the AlertMeter® via tablets and their smart phones. They also completed the PVT on their tablets. We assessed WMTB at three time points: (a) One hour of wakefulness (b) 17 hours of wakefulness, (c) 34 hours of wakefulness. Participants were provided meals and snacks without caffeine and limited sugars. They were able to play games in the room, take monitored walks, and watch television. They were not permitted to use any stimulants. During the wakefulness phase of the study, graduate research assistants monitored the participants. Additionally, Drs. Owen, Sherry and graduate research assistants monitored the participants. At the end of the 36 hours, participants were given a room at the hotel to sleep at least eight hours. They were assessed the following morning and asked to continue to wear the Fitbit for two more days after the wakefulness portion of the study.

3. RESULTS

Table 3.1 displays the overall means (M) and standard deviations (SD) of reaction times for the Mobile and Web/Tablet based AlertMeter® and the PVT and KSS by study participant. The empty cells indicate missing data due to instrument malfunction. In the analyses, please note that the PVT, Lapses and KSS means are the same for both the AlertMeter® Mobile and Web, which were taken at the same time the PVT. In addition, it is also evident from the inspection of Table 3.1 that PVT data was not obtained for participants 12 and 15. These participants did complete the AlertMeter® but provided insufficient data from the PVT to make meaningful comparisons. However, their AM scores were included in the growth models, as maximum likelihood estimation can be inclusive of missing data.

This table also reflects the reexamination of the data, to adjust and align the PVT data to be in line with actual local time. The data were initially reported using Coordinated Universal Time or UTC, the successor to Greenwich Mean Time (GMT). Denver is UTC-7 or seven hours behind UTC-0 (London, UK). Realigning the data decreased the variability in the observations and permitted synchronization of the means across AlertMeter® and PVT administrations.

Table 3.1 Descriptive Statistics for AlertMeter® RT, Mobile and Web by Participant

		AM - RT		PVT-RT		Lapses		KSS	
ID	Test	Mean	SD	Mean	SD	Mean	SD	Mean	SD
D02	Mobile	2712.63	1268.96	326.68	28.94	10.40	4.59	4.81	2.32
	Web	1993.14	979.01	327.12	29.29	10.64	4.99	4.71	2.20
D03	Mobile	2283.65	1025.23	367.22	130.23	11.69	9.11	5.15	1.76
	Web	1923.66	848.76	363.25	126.36	10.93	8.83	5.13	1.95
D04	Mobile	2233.59	1004.89	342.98	92.51	10.81	6.99	4.88	1.66
	Web	2438.14	1354.27	339.17	87.93	10.94	6.63	4.78	1.59
D05	Mobile	2524.31	1302.46	553.88	124.58	29.18	7.73	5.83	1.35
	Web	2068.15	1323.13	553.88	124.58	29.18	7.73	5.83	1.35
D06	Mobile	2438.89	1190.76	378.85	62.46	17.18	7.03	4.29	1.88
	Web	2473.31	1324.07	378.85	62.46	17.18	7.03	4.29	1.88
D07	Mobile	2110.83	724.81	291.53	22.56	6.67	2.72	5.50	2.10
	Web	1640.38	677.88	300.47	26.65	7.50	3.42	6.00	1.85
D08	Mobile	1835.42	653.18	347.98	51.50	13.31	6.84	5.87	1.63
	Web	1588.29	669.77	347.98	51.50	13.31	6.84	5.87	1.63
D09	Mobile	2258.74	1089.54	483.96	164.26	17.14	7.81	4.86	1.66
	Web	2367.43	1221.19	481.97	163.47	16.86	7.82	4.86	1.64
D10	Mobile	1611.89	688.59	320.80	31.48	9.75	7.03	3.77	1.73
	Web	1641.6	905.26	320.80	31.48	9.75	7.03	3.77	1.73
D11	Mobile	1579.86	630.75	337.22	38.96	10.53	5.37	5.00	1.38
	Web	1567.59	742.84	337.22	38.96	10.53	5.37	5.00	1.38
D12	Mobile	1593.71	726.00						
	Web	1671.46	819.29						
D13	Mobile	1900.78	710.50	322.05	31.77	11.06	7.06	3.44	1.46
	Web	1970.93	1015.28	322.68	31.57	11.35	7.02	3.54	1.50
D14	Mobile	1929.37	773.40	331.60	49.48	12.50	7.63	4.38	1.61
	Web	1891.94	840.89	331.48	49.33	12.50	7.61	4.39	1.61
D15	Mobile	1623.08	757.66			21.50	13.44		
	Web	1566.72	726.73			21.50	13.44		
D16	Mobile	1475.69	740.56	306.81	39.48	21.50	13.44		
	Web	1388.51	633.65	306.81	39.48	7.80	6.99	4.35	2.12
D17	Mobile	1743.06	968.28	456.46	131.82	7.80	6.99	4.35	2.12
	Web	1645.05	748.18	456.46	131.82	17.61	4.36	4.44	1.35
D18	Mobile	1262.94	460.66	383.48	37.19	17.61	4.36	4.44	1.35
	Web	1443.42	672.90	383.48	37.19	18.35	4.93	3.67	1.87

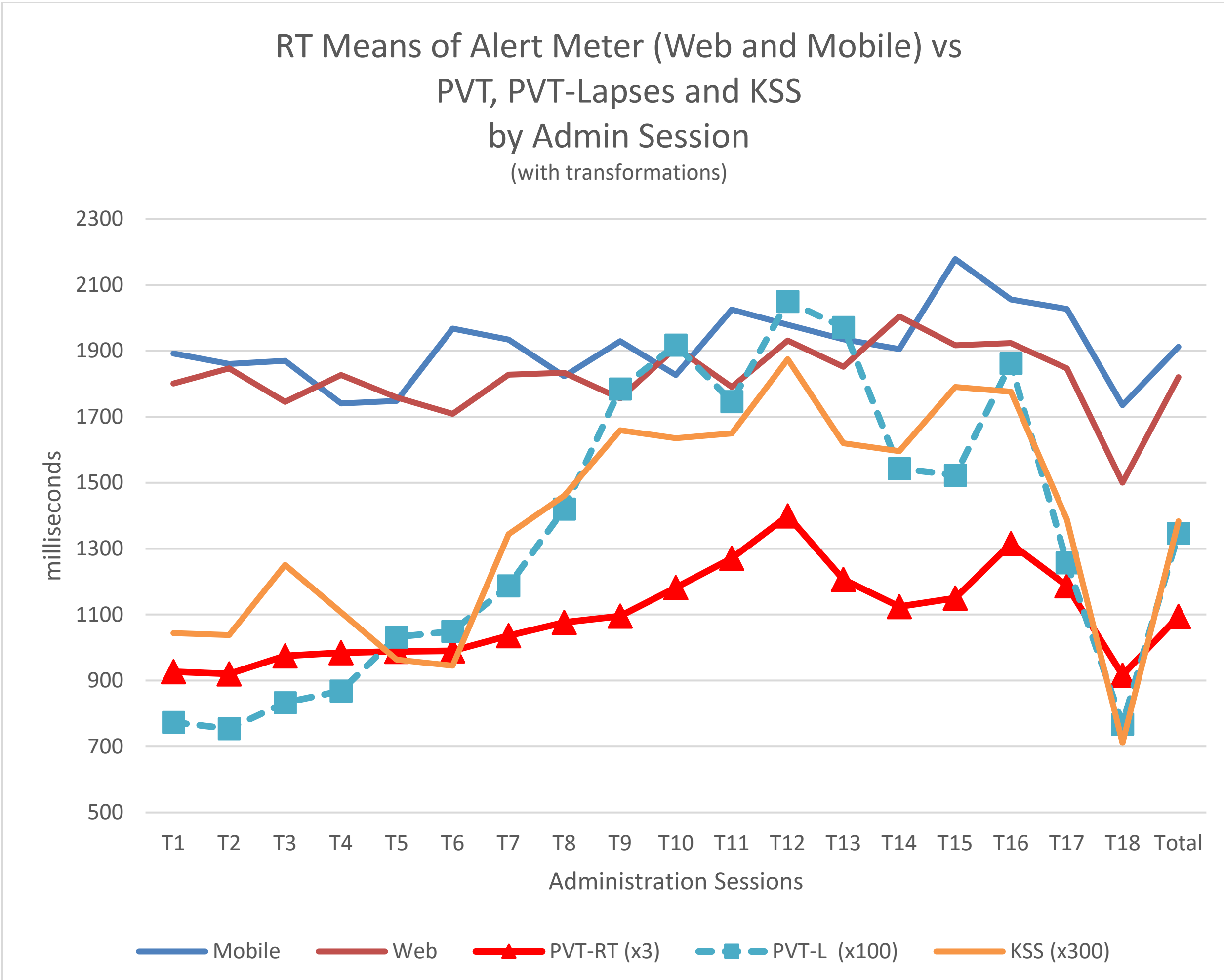

Figure 3.1 Means of AM-RT, PVT and KSS by Administration session

Table 3.2 Descriptive Statistics for AlertMeter® RT, Mobile with Memory by Participant

ID	AM Mob RT	SD	PVT RT	SD	PVT Lapses	SD	KSS Self Report	SD	Score	SD
D2	2424.62	1072.061	327.12	29.19	10.64	4.97	4.71	2.19	0.29	0.05
D3	2171.06	1132.537	364.3	126.41	11.00	8.77	5.17	1.95	0.41	0.10
D4	2379.25	1269.406	339.17	87.72	10.94	6.61	4.78	1.58	0.33	0.11
D5	2153.54	1122.431	553.88	124.04	29.18	7.70	5.83	1.35	0.40	0.09
D6	2251.28	1192.341	378.85	62.22	17.18	7.00	4.29	1.87	0.36	0.09
D7	1772.29	813.94	300.48	26.51	7.50	3.40	6.00	1.83	0.50	0.06
D8	1793.9	766.686	347.98	51.33	13.31	6.82	5.87	1.63	0.47	0.03
D9	1909.89	1113.428	475.43	167.33	16.14	8.15	5.00	1.61	0.42	0.09
D10	1547.98	726.347	320.8	31.36	9.75	7.00	3.77	1.72	0.55	0.07
D11	1460.93	618.516	337.22	38.86	10.53	5.35	5.00	1.37	0.62	0.06
D12	1548.98	744.745							0.55	0.10
D13	1821.64	891.807	322.76	31.47	11.39	6.99	3.56	1.50	0.46	0.05
D14	2004.81	945.465	331.48	49.21	12.50	7.59	4.39	1.61	0.39	0.05
D15	1448.31	592.7							0.59	0.10
D16	1422	661.481	306.81	39.36	7.80	6.97	4.35	2.12	0.55	0.13
D17	1551.21	846.527	456.46	131.50	17.61	4.35	4.44	1.34	0.51	0.09
D18	1323.72	627.26	383.48	37.10	18.35	4.92	3.67	1.86	0.63	0.09
Total	1814.62	974.173	364.02	99.83	13.36	8.20	4.62	1.86	0.47	0.13

Table 3.3 Descriptive Statistics for AlertMeter® RT, Web/Tablet with Memory by Participant

	AM Web - RT		AM Score		PVT-RT		PVT-Lapse		KSS	
ID	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
D2	2031.28	1141.10	0.41	0.06	327.12	29.19	10.64	4.97	4.71	2.19
D3	1984.77	1072.50	0.42	0.13	364.30	126.41	11.00	8.77	5.17	1.95
D4	2415.46	1380.81	0.35	0.08	339.17	87.72	10.94	6.61	4.78	1.59
D5	2108.92	1181.43	0.40	0.07	553.88	124.04	29.18	7.70	5.83	1.35
D6	2374.93	1320.24	0.34	0.07	378.85	62.22	17.18	7.00	4.29	1.87
D7	1622.29	770.21	0.52	0.07	300.48	26.51	7.50	3.40	6.00	1.83
D8	1622.00	684.78	0.47	0.09	347.98	51.33	13.31	6.82	5.87	1.63
D9	1976.56	1165.70	0.41	0.09	475.43	167.33	16.14	8.15	5.00	1.61
D10	1686.05	873.18	0.51	0.05	320.80	31.36	9.75	7.00	3.77	1.72
D11	1426.87	737.45	0.58	0.05	337.22	38.86	10.53	5.35	5.00	1.37
D12	1611.43	940.48	0.50	0.14						
D13	1917.41	1019.08	0.45	0.03	322.76	31.47	11.39	6.99	3.56	1.50
D14	1884.36	918.88	0.44	0.05	331.48	49.21	12.50	7.59	4.39	1.61
D15	1589.61	808.50	0.50	0.07						
D16	1438.89	627.81	0.57	0.10	306.81	39.36	7.80	6.97	4.35	2.12
D17	1678.51	861.20	0.47	0.05	456.46	131.50	17.61	4.35	4.44	1.34
D18	1462.82	725.60	0.56	0.05	383.48	37.10	18.35	4.92	3.67	1.86
Total	1808.77	1018.45	0.47	0.11	364.02	99.83	13.36	8.20	4.62	1.86

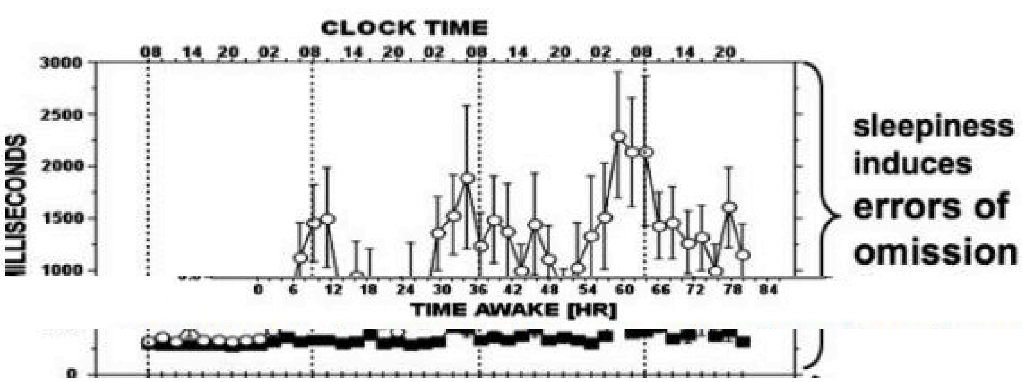

Figure 3.2 PVT results from Lim & Dinges (2008)

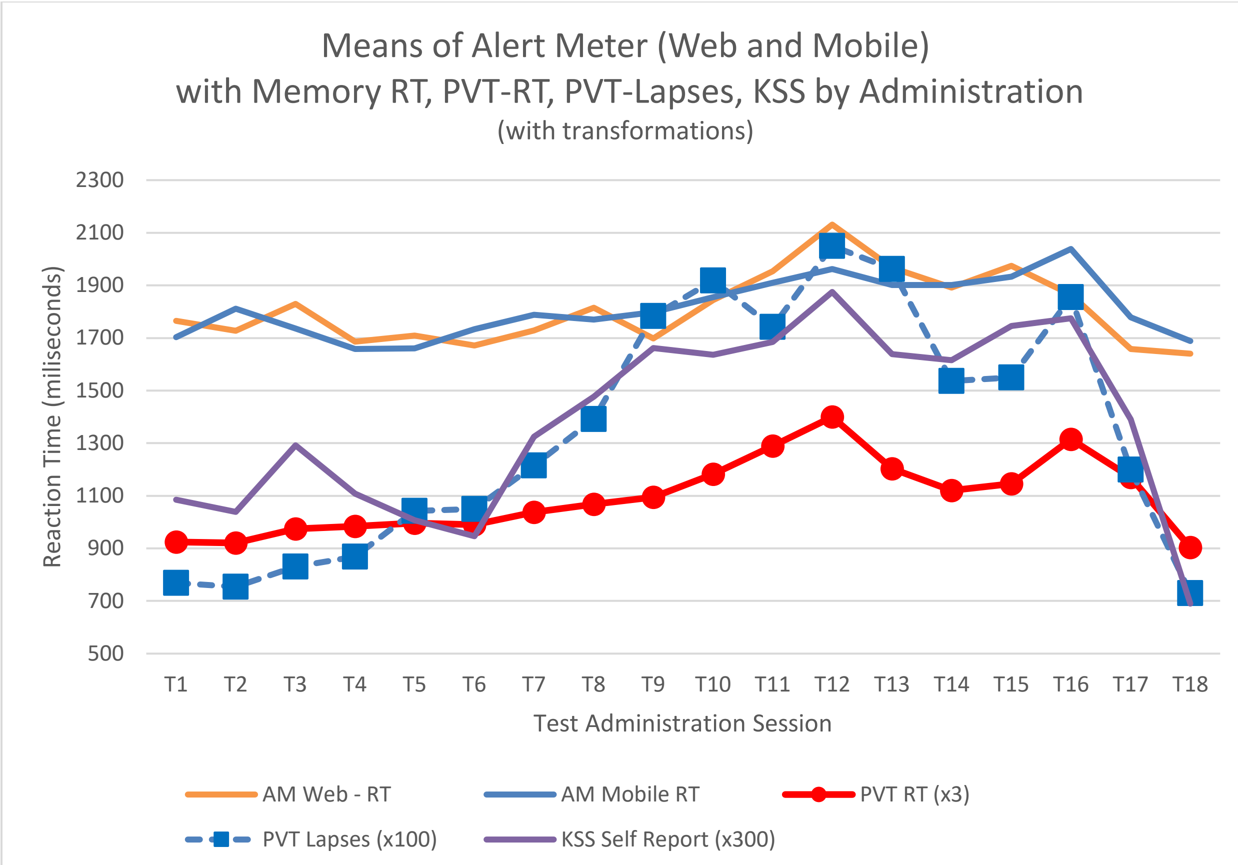

Figure 1.3 AlertMeter® Memory AM-RT, PVT-RT, PVT-Lapses, and KSS by Admin sessions

3.1 Individual Performance on PVT

The sample of study participants completed the assessment according to the predefined schedule. Results of the 18 assessments on all measures during the 36-hour extended wakefulness portion of the study, between 8 a.m. Friday and 8 p.m. Saturday provided a description of the alertness of the participants, the final assessment was conducted after the participants had completed 12 hours of time in bed (Sunday between 8 a.m. and 10 a.m.). Study participants whose performance on the PVT deteriorated over the course of the study are presented in Figure 3.4. Note that there is a marked increase in reaction time during the testing administrations after session #9, or 2 a.m., on Saturday morning, continuing until about noon on Saturday, as would be expected. The participants not plotted did not show much sensitivity to the effects of extended wakefulness and did not noticeably increase their reaction times over the course of the study.

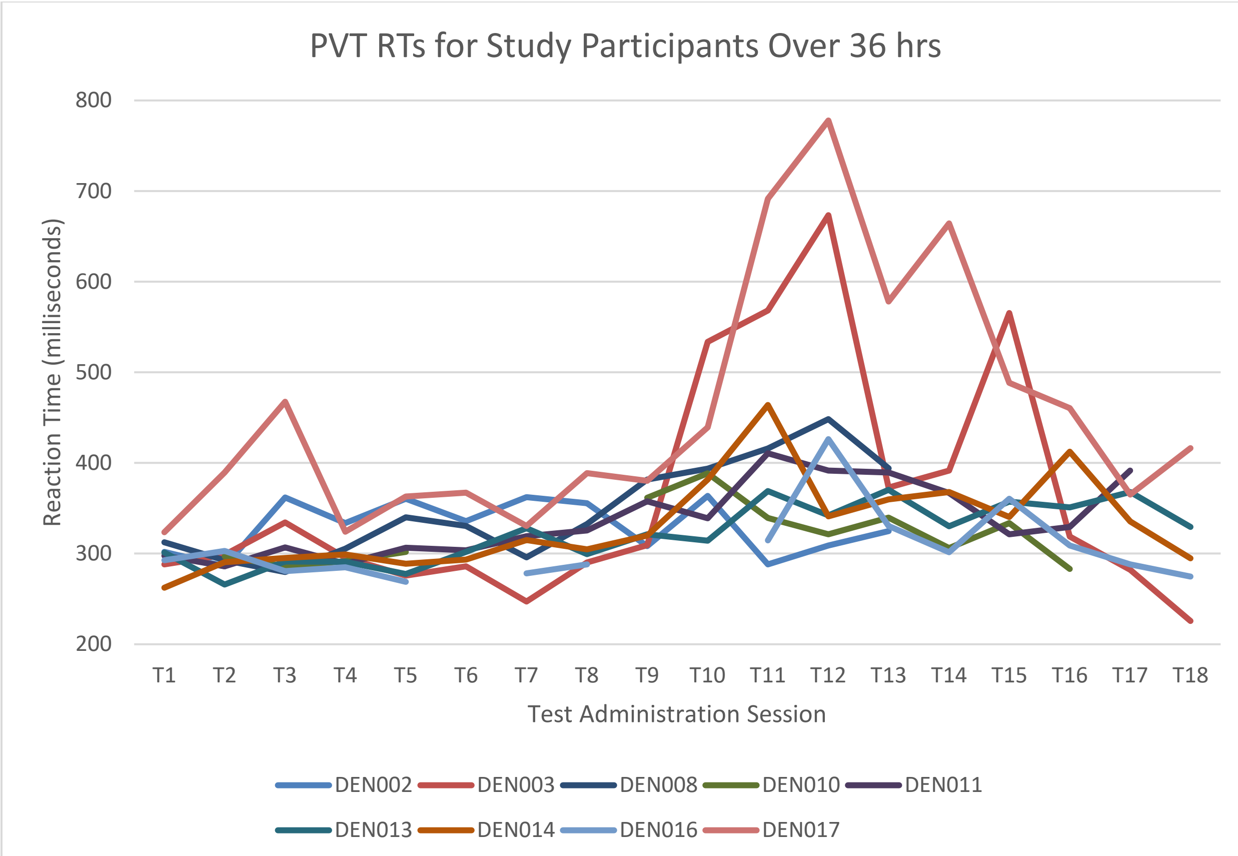

Figure 3.4 PVT-RTs for study participants over 36 hours

The results in Figure 3.4 also show that at 22 hours of wakefulness 58% of the participants had average PVT-RTs that were at least 20% greater than their 10 a.m. Friday morning baseline. Further inspection reveals that fully 83% demonstrated decreased performance on the PVT, or increased reaction times after 24–26 hours of wakefulness. Improving one's performance means faster performance and that they were essentially more resistant to the effects of fatigue and therefore suffered less cognitive decline being able to maintain vigilance and reaction time.

The results of the AM-Web-RTs in Figure 3.5 show that at 22 hours of wakefulness, 47% of the study participants were demonstrating noticeable fatigue with increase reaction times relative to their 10 a.m. Friday morning baseline. Only 24% were demonstrating increased reaction times after 24 hours of wakefulness. The participants' performance on the AM-Web was at its poorest after 16 hours of wakefulness with 58% demonstrating increased reaction times.

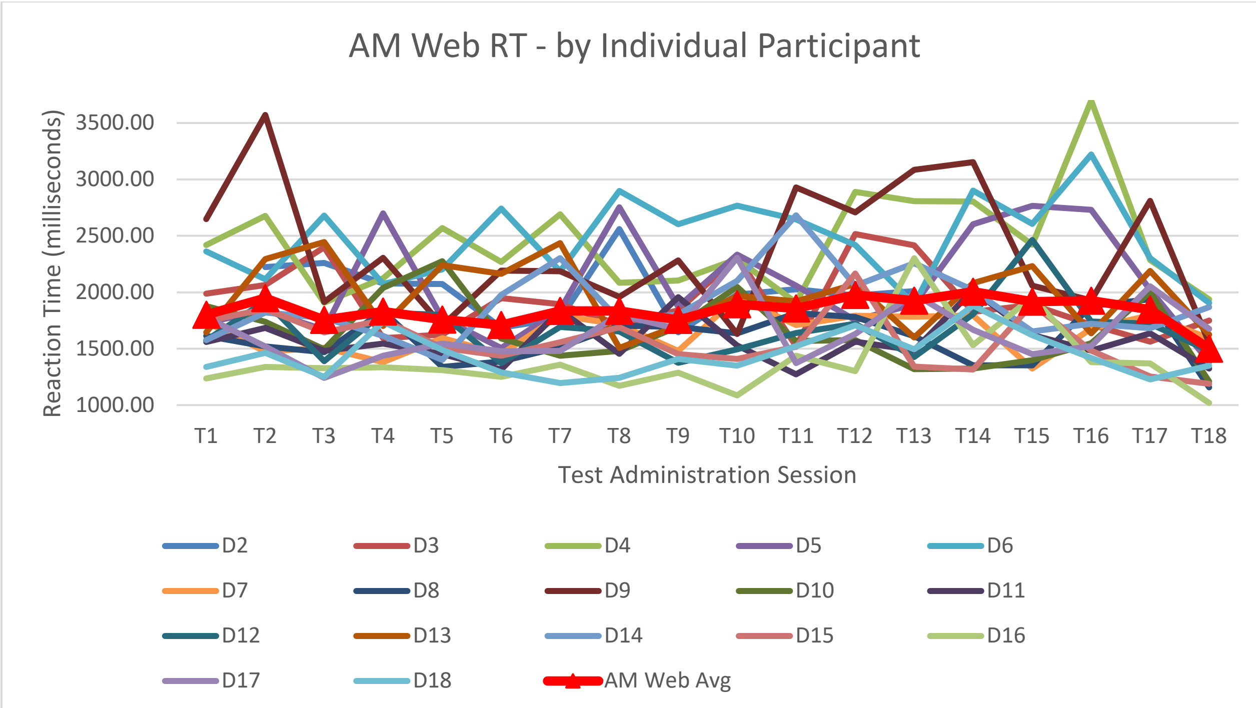

Figure 3.5 AM Web-RT by Individual

3.2 Main Results

To test the association between PVT and AlertMeter®, we conducted a multilevel growth curve model where time (repeated measures) was nested within individuals. This approach to growth curve modeling is superior to the traditional repeated measures analyses. That is, we were better able to account for the interdependencies in the data when the same individuals complete assessments multiple times during a study. For example, some participants may be more resistant to fatigue, which is accounted for to some extent by the within subjects nested design of this analysis. To account for predicted fluctuations in the scores over time we modeled a linear, quadratic, and cubic change functions. We modeled the random effects for participant variability around the intercept (e.g., participants might vary in their AM/PVT scores) and the participant variability in change over time. This latter part allows a better estimation as it is clear that the participants had different trajectories of change (see Figure 3.5).

3.2.1 Results for AlertMeter®-Mobile, No Memory

The results for the AlertMeter® (non-memory) and PVT are presented in Table 3.4. These results demonstrate that AlertMeter®-non memory tests are sensitive to change over time consistent with the likely increase in fatigue associated with extended wakefulness. Additionally, the AlertMeter® sensitivity to a slight recovery, followed by a performance decline, at T14 and T15 (increase in RT) for both Mobile and Web/Tablet versions again supports the AlertMeter® as a valid measure of fatigue. The difference between the AM and PVT is likely due to the more complex cognitive demands required to perform the AM tasks as opposed to the simpler vigilance task required on PVT. Moreover, consistent with a concurrent validity study, the patterns were significant and similar to the pattern of change demonstrated by the PVT (overall correlation between the PVT and AM-mobile, and AM-web in the growth curves was $r = .62$, .61, respectively). Lack of significance of a linear model of change for the PVT is obvious from the shape of the curves presented in Figures 3.3 and 3.6, as compared to the Lim & Dinges (2008) results in Figure 3.2, and is consistent with previous research that has found a non-linear pattern for PVT-RT performance over extended periods of wakefulness, suggesting that a simple linear model of fatigue or alertness detection is not as sensitive across all assessment techniques. The present results are also consistent with Lim & Dinges (2008) findings with respect to lapses after 20 hours of wakefulness (Dinges, 2008). The pattern of results suggests both a linear and a cubic decline in performance as hours of wakefulness increases (less sleep, more fatigue, poorer performance).

Table 3.4 AlertMeter® (non-memory) and PVT Growth Models Results

	AM-Mobile	AM-Web	PVT
	Coefficient (SE)	*Coefficient (SE)*	*Coefficient (SE)*
Baseline	1962.92 (119.86)***	1803.11 (94.82)***	329.84 (18.65)***
Linear Growth	-75.32 (28.28)*	-74.13 (28.83)*	-14.27 (8.62)
Quadratic Growth	12.84 (3.90)**	15.22 (3.99)**	4.20 (1.19)**
Cubic Growth	-0.51 (0.15)**	-0.67 (0.15)***	-0.19 (0.04)***

Notes. *p < .05, **p < .01, ***p < .001

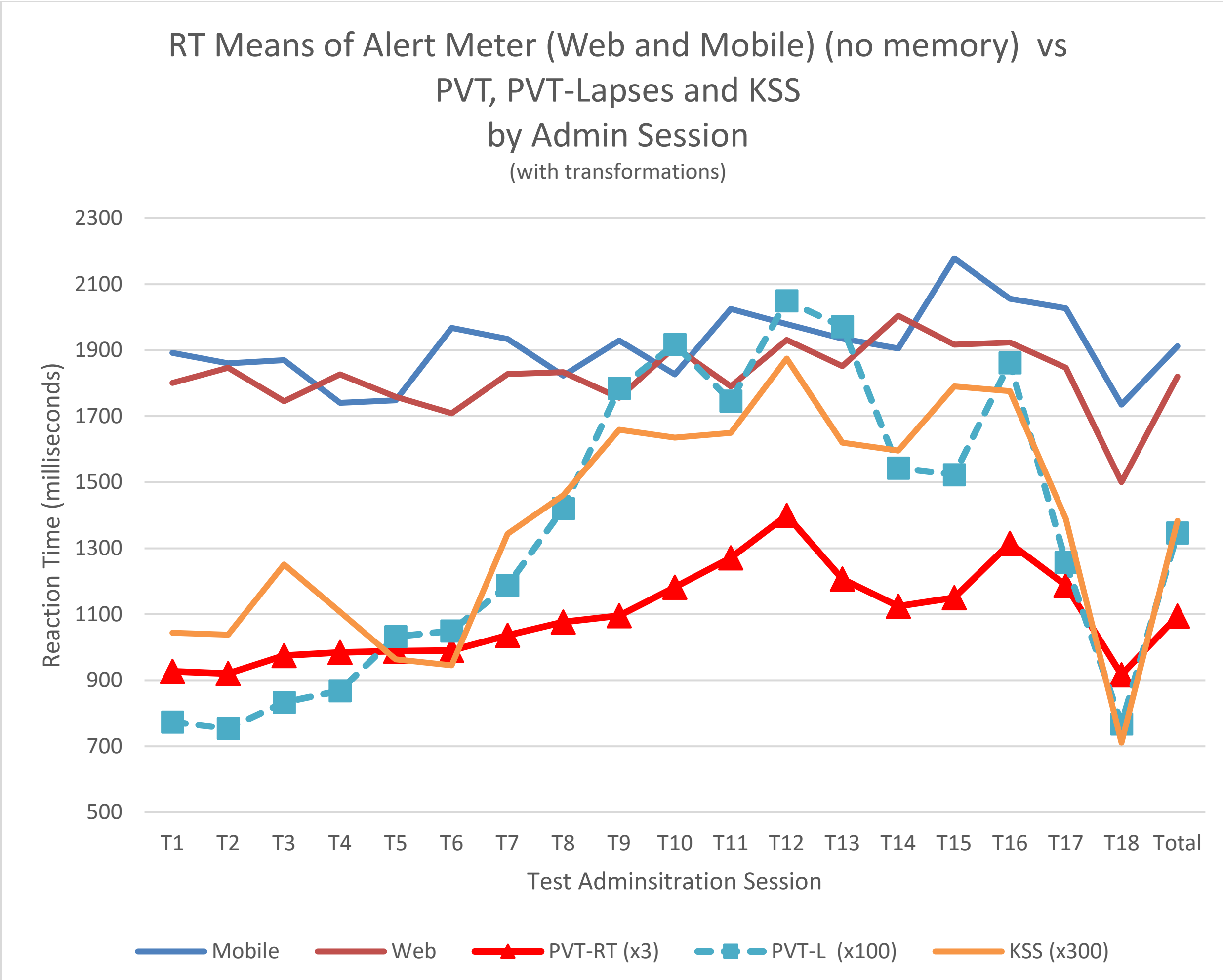

Figure 3.6 AlertMeter®-Mobile (non-memory) RT - across tessing sessions

Note: *PVT, Lapses and KSS were transformed to permit ease of comparison in the figure. The actual values of the KSS and Lapses were not in milliseconds.*

3.2.2 Results for AlertMeter®-Mobile, with Memory

Next, we conducted the same growth curve analyses on AlertMeter® with memory. As seen in Table 3.5, the results again depict significant changes in reaction times for both Mobile and Web versions of AlertMeter® over time consistent with extended wakefulness. Incidentally, the overall correlation between the PVT and AM-mobile, and PVT and AM-web growth curves was $r = .61$, .60, respectively). The change patterns are consistent with and similar to the non-memory versions of the AlertMeter®, with one exception—AlertMeter®-mobile memory did not demonstrate a significant linear change; however, other parameters were significant. The lack of significant linear change is consistent with the PVT. The overall trajectories of change for AlertMeter® with and without memory are illustrated in Figures 3.5 and 3.6. Overall, these results support the use of AlertMeter® as a valid measure of fatigue.

Table 3.5 AlertMeter® with Memory and PVT Growth Models Results

	AM-Mobile *Coefficient (SE)*	AM-Web *Coefficient (SE)*
Baseline	1710.76 (113.45)***	1873.11 (81.99)***
Linear Growth	-39.36 (34.19)	-101.83 (27.04)**
Quadratic Growth	12.22 (4.22)**	19.21 (3.70)**
Cubic Growth	-0.57 (0.16)**	-0.83 (0.15)***

Notes. *p < .05, **p < .01, ***p < .001

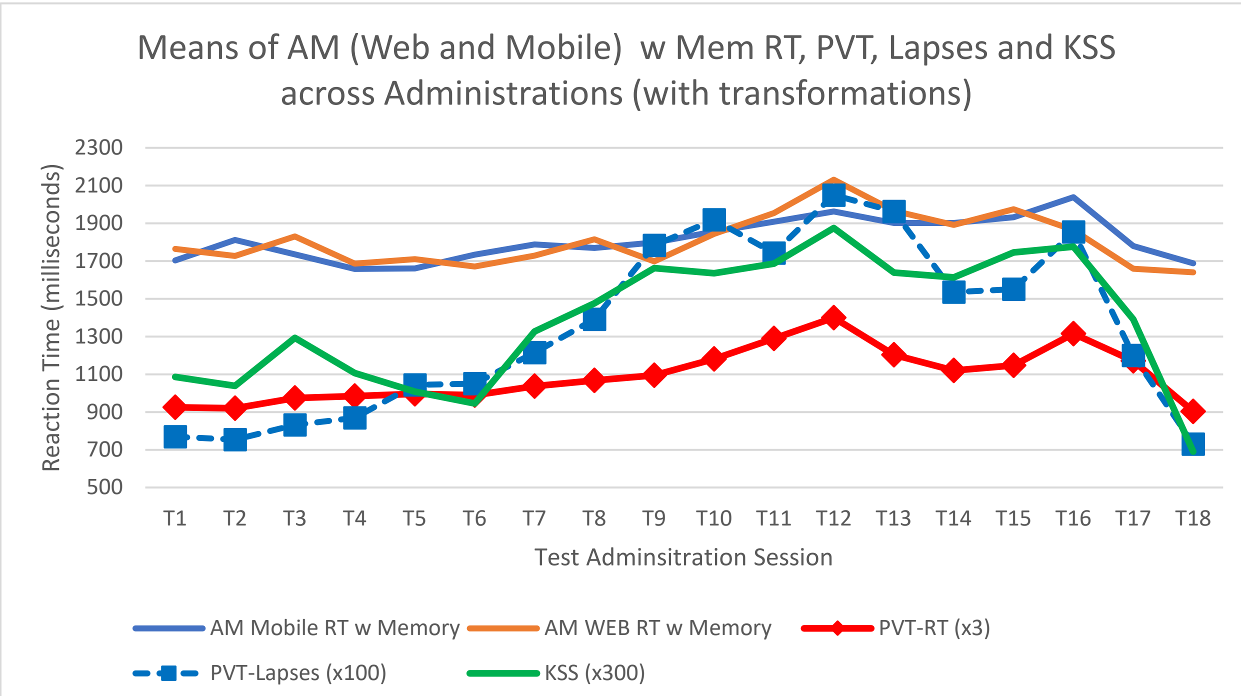

Figure 3.7 Means of AM Mobile and Web PVT-RT, PVT-Lapses and KSS

Table 3.6 AlertMeter® and KSS & PVT Lapses

	AM-Mobile (nm) *r (p-value)*	AM-Web (nm) *r (p-value)*	AM-Mobile (m) *r (p-value)*	AM-Web (m) *r (p-value)*
KSS	.29 (< .001)	.31 (< .001)	.31 (< .001)	.19 (.006)
Lapses	.16 (.03)	.29 (< .001)	.22 (.002)	.24 (< .001)

We also correlated AlertMeter® with participants' self-report of sleepiness (KSS) and the number of lapses from the PVT. There were significant correlations for all four versions of AlertMeter® (see Table 6). All correlations were statistically significant, with typically medium-sized effects. These findings provide further support for the validity of the AlertMeter® in detecting perceived fatigue and the performance decrements associated with extended wakefulness.

3.2.3 Results for AlertMeter®-Mobile, No Memory by Bins

The mean reaction times of the bins for the AlertMeter® Mobile (no Memory) were plotted in Figure 3.8 for ease of inspection. As can be seen the spike after T11 – 6 a.m. Saturday reflects the effect of extended hours of wakefulness and demonstrates validity of the AlertMeter®. The statistically significant correlations between AM and PVT also demonstrate concurrent validity with the PVT a standard measure of fatigue. The fact that the AM RTs are also correlated with the KSS scores, which are self-report subjective measures of fatigue also provides substantial evidence of face validity.

As seen in Figure 3.8 and Table 3.7, only Bin 202 was non-significant across all three measures (PVT, Lapses, and KSS). Bins 102, 202 and 203 did not correlate with PVT-RT or PVT-Lapses but, with the exception of Bin #202, did correlate with the KSS. The shape of the curve for Bins 202 and 203 are relatively flat, consistent with a low correlation with the PVT. Accordingly, we suggest further refinement of Bins 102, 202 and 203. The remaining bins were significantly correlated with the PVT and thus operate as valid measures of fatigue. However, bins 102, 201, 202, and 203 did not correlate with PVT-Lapses and may also require further refinement.

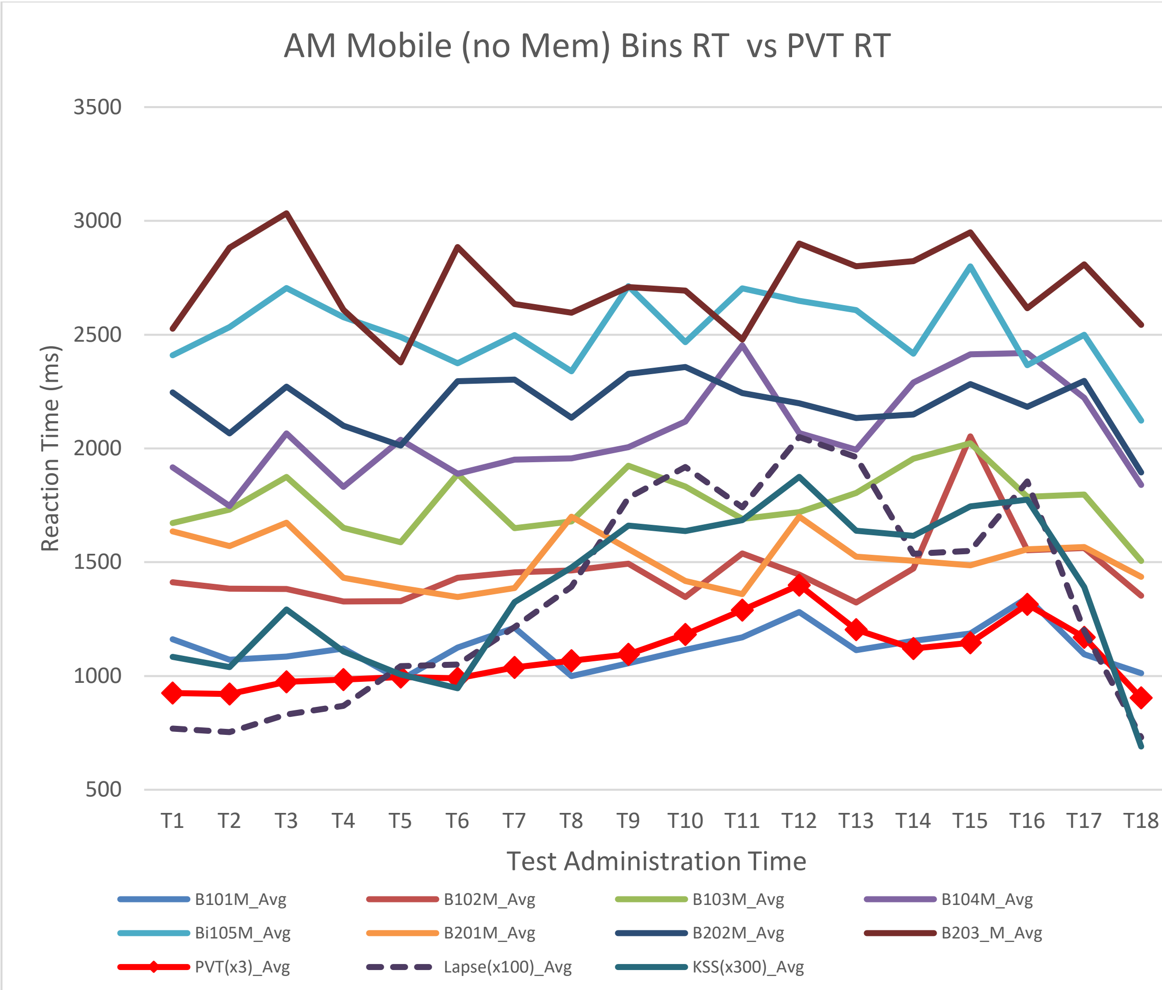

Figure 3.8 AM Mobile (nm) RT w Bins

Table 3.7 Correlations of AM Mobile RT (nm) with PVT and KSS by Bin

Correlation between AM Mobile (no mem) RT and PVT, Lapses and KSS by Bin.			
	PVT-RT	PVT-Lapses	KSS
101	.316**	.227**	.162**
102	0.00	-0.01	.124*
103	.181**	.186**	.254**
104	.204**	.168**	.198**
105	.136*	.127*	.181**
201	.153*	0.10	.140*
202	0.11	-0.02	0.04
203	0.05	0.06	.156*
** Correlation is significant at the 0.01 level (1-tailed).			
* Correlation is significant at the 0.05 level (1-tailed).			

3.2.4 Results for AlertMeter®-Web/Tablet, No Memory

The mean reaction times of the bins for the AlertMeter® Table/Web (no Memory) were plotted in Figure 3.9 for ease of inspection. Again, the spikes at T12 and T13 – 8 a.m. and 10 a.m. Saturday reflect the effect of extended hours of wakefulness and demonstrates the validity of the AlertMeter®. The presence of statistically significant correlations between AM and PVT also demonstrates concurrent validity with the PVT a standard measure of fatigue. The fact that the AM RTs are also correlated with the KSS scores indicates the presence of substantial face validity.

As seen in Figure 3.9 and Table 3.8, Bins 101, 102, 103, 105, and 203 showed significant associations with the PVT and demonstrate validity as measures of fatigue. Bins 103, 202, and 203 also show good correspondence with the KSS and PVT-Lapses. The shape of the curve for Bins 104, 201, and 202 are relatively flat, with little change in their reaction times, which is consistent with a low and non-significant correlation with the PVT. The other bins were not significantly associated with the PVT. Accordingly, 104, 201, and 202 may need improvement.

It should be noted that the difference in magnitude between the RT for the AM Bins and the RT for the PVT reflects the underlying difficulty of the items in the bins in comparison to the relatively quick responses to the PVT stimuli. While the curves graphed in Figure 9 have been subjected to a linear transformation, the fact remains that PVT reaction time is one-third the magnitude of the AM reaction time. This suggests that the AM task requires more mental resources beyond that of the simple recognition of the presence of a stimulus. It draws on memory, image comparison and stimulus recognition and therefore mirrors the additional cognitive processes required in the workplace.

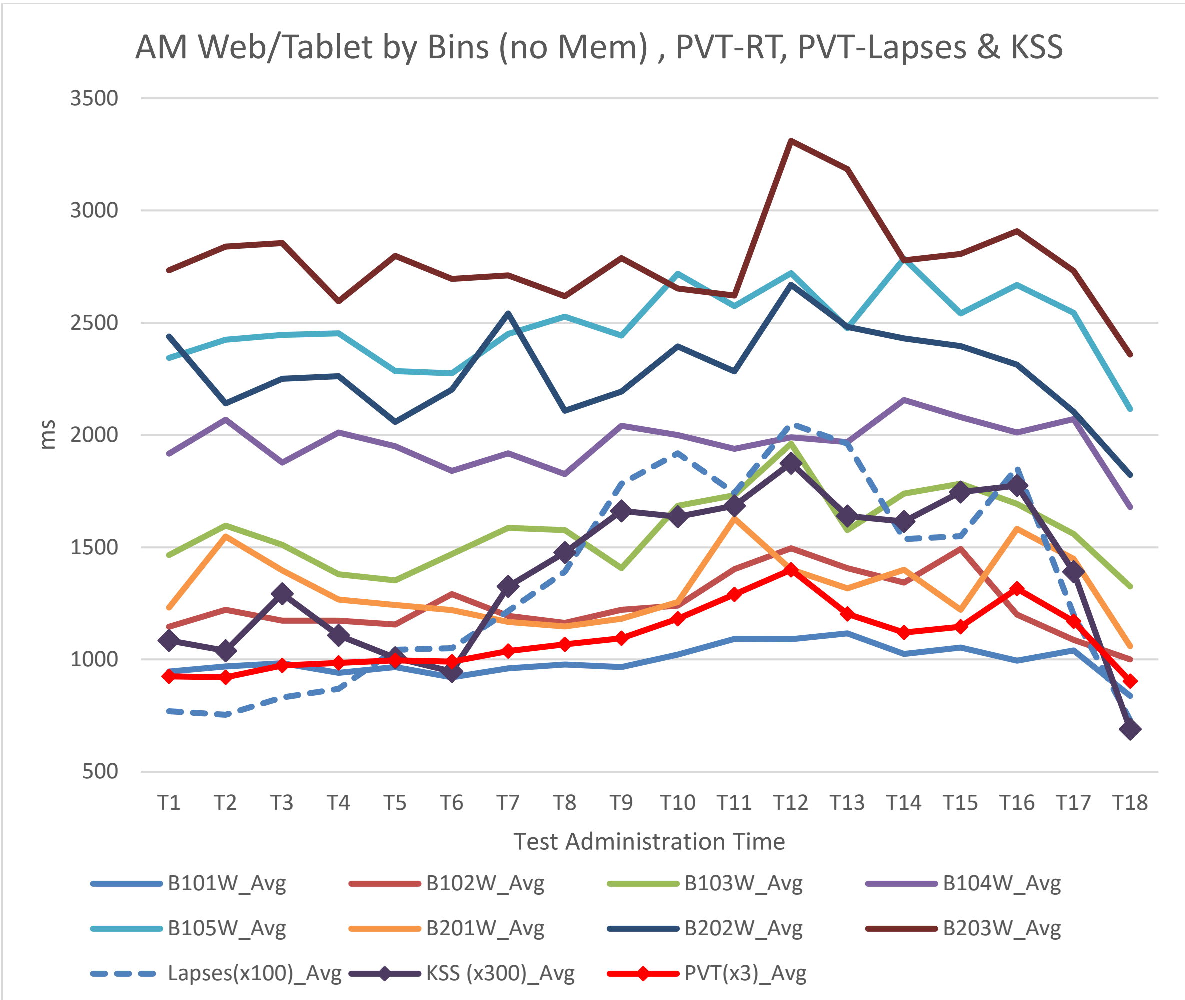

Figure 3.9 Mean AM RTs for Web/Table (nm) by Bins, PVT, Lapses and KSS

Table 3.8 Correlations AM Web/Tablet RT (nm) with PVT-RT, PVT-Lapses KSS by Bin

Correlation AM Mobile RT and PVT, Lapses and KSS by Bin

Bin	AM Web PVT-RT	PVT-Lapses	KSS
101	.251**	.235**	.129*
102	.165**	.147*	.077
103	.148*	.081	.155*
104	.055	.057	.031
105	.254**	.233**	.232**
201	.011	.021	.100
202	.099	.107	.072
203	.242**	.251**	.063

** Correlation is significant at the 0.01 level (1-tailed).
* Correlation is significant at the 0.05 level (1-tailed).

3.2.5 Results for AlertMeter®-Mobile, with Memory by Bins

The graph in Figure 3.10 and Table 3.9 represents the average reaction time response at each two-hour testing of the study participants by bins for the AlertMeter® Mobile with memory. Only Bin 102, 202, and 301 did not correlate significantly with the PVT-RT measure. Bin 301, which was not significant, appears as a relatively flat line, suggesting little differentiation across the 18 testing sessions and therefore, not likely sensitive to the presence of fatigue in the study participants.

Correlations between the AM Mobile RT with memory bins with PVT-Lapses shows significant correlations for Bins 101, 102, 103, 104 and 105, 203 and 302 and 303. Bins 301 and 201 were not significant. Thus, there is a similar pattern of results with errors of omission (lapses) demonstrated with the bins results.

Last, there is a similar pattern of results of correlations between the bins and the KSS self-reported subjective alertness and fatigue as measures by the KSS. Only Bin 301 and 302 were not significantly correlated with the KSS. Interesting, linty the average correlation between the AM Mobile RT and the KSS was higher than the correlation between the AM Mobile with memory RT and the PVT-RT and PVT-Lapses, suggesting that there is a closer match between the AM Mobile with memory and subjective assessments of fatigue. To note, the correlations for any given bin will naturally be lower than the correlations between the growth curves.

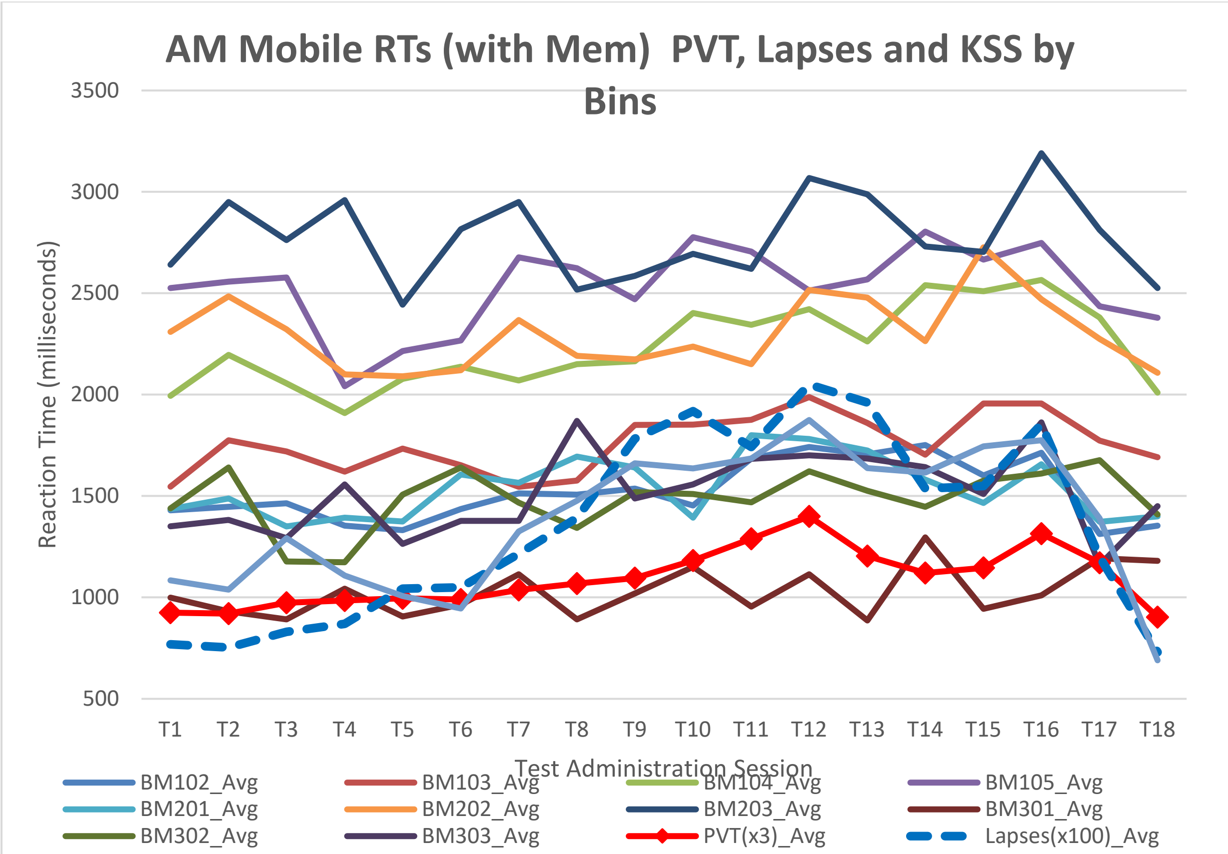

Figure 3.10 Mean RTs of bins for AM-Mobile - Memory and PVT

Table 3.9 Correlations AM Mobile RT (with Mem) PVT-RT, PVT-Lapses, KSS by Bin

Bin	PVT-RT	PVT-Lapses	KSS
101	.174**	.170**	.226**
102	.063	.110**	.190**
103	.151**	.138**	.166**
104	.199**	.119**	.203**
105	.094**	.117**	.221**
201	.116*	.029	.171**
202	.050	.012	.160**
203	.103*	.100*	.122**
301	-.009	.033	.076
302	.097*	.138**	.076
303	.177**	.207**	.163**

** p< 0.01 level (1-tailed), * p< 0.05 level (1-tailed).

3.2.6 Results for AlertMeter®-Web/Tablet, with Memory, by Bins

Results of the analyses for AlertMeter® Web/Tablet with Memory are presented in Figure 3.11 and Table 3.10. As can be seen from the graphic depiction of the reaction times performance is similar to previous results and demonstrates that the AlertMeter® is detecting the decline in performance and the increase in reaction times as the length of wakefulness increases. Similarly, all bins, except Bin 301 were significantly correlated with PVT-RT indicting that the bins are also sensitive to changes in reaction time. Bins 102, 201, and 301 were not correlated significantly with PVT-Lapses. Bins 105, 203, and 301 were not significantly correlated with the KSS subjective ratings of fatigue. In comparison, this set of bins on the AM Web/tablet appears to be doing quite well at detecting fatigue as measured by the PVT and subjective self-assessments of fatigue.

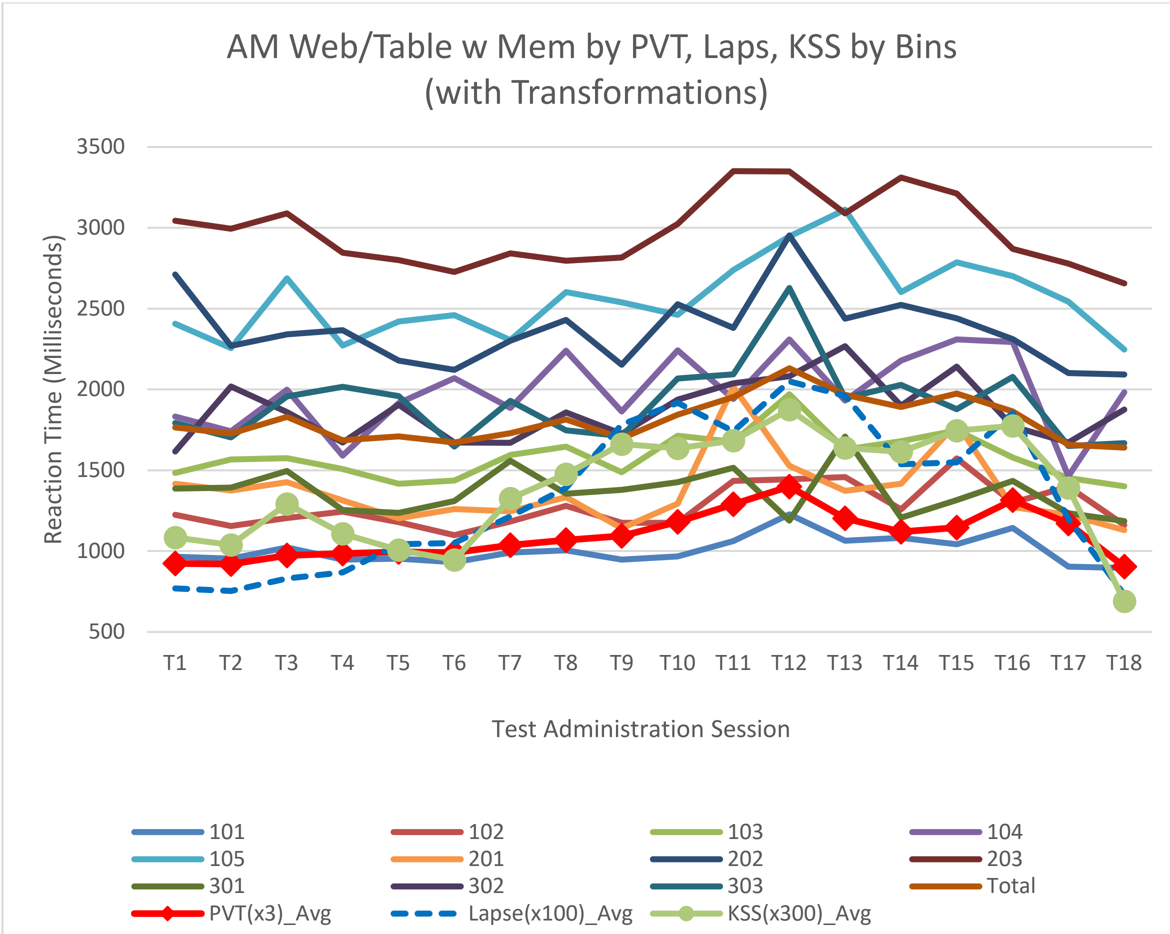

Figure 3.11 Mean RT of significant bins for AM-Tablet - w Memory and PVT

Table 3.10 Correlations AM Web/Tablet RT (with mem) PVT-RT, PVT-Lapse, KSS by Bin

Bin	PVT-RT	PVT-Lapses	KSS
101	.174**	.182**	.104**
102	.091*	0.063	.121**
103	.124**	.098**	.082**
104	.135**	.137**	.161**
105	.145**	.132**	0.058
201	.159*	0.062	.201**
202	.129**	.080*	.078*
203	.147**	.145**	0.021
301	-0.008	0.025	0.106
302	.152**	.152**	.116**
303	.257**	.226**	.107*

** Correlation is significant at the 0.01 level (1-tailed).
* Correlation is significant at the 0.05 level (1-tailed).

On inspection, Bins 101, 103, 105, 201, 203, and 303 look as if they are the most sensitive to fatigue due to the fact that they have an increasing slope during the early morning hours of Saturday, peaking at T12- 8 a.m. on Saturday and then returning to a more optimal level.

3.3 AlertMeter® Memory Correlations with WMTB Tests

We correlated the AlertMeter® Memory Accuracy scores with the Working Memory Test Battery (WMTB). The specific AM test will provide the best one-to-one comparison with the WMTB. We collected data at three time points for the WMTB. In the Table 3.11, we display the average correlations, which were non-significant, for each time point with the baseline or 10 a.m. Friday testing. Results of the analysis revealed no clearly discernable pattern in the correlations, with some in the positive direction and others in the opposite direction. There does not appear to be a significant correlation between the AM Memory Accuracy test and memory accuracy measured at the three different time points using standard measures of memory.

Table 3.11 Average correlations between AlertMeter® and WMTB

AlertMeter®	**WMTB Average Correlations**
	r
Time 1	
Mem. Accuracy-Mobile	.22
Mem. Accuracy-Tablet	.22
Time 2	
Mem. Accuracy-Mobile	.23
Mem. Accuracy-Tablet	.21
Time 3	
Mem. Accuracy-Mobile	.25
Mem. Accuracy-Tablet	.19

Further inspection of the 18 AlertMeter® Mobile Shape Pair Recognition and Web Shape Pair Recognition memory accuracy scores revealed a fairly flat profile suggesting that the memory test was not actually differentiating or varying much across time. The memory accuracy score ranges from 1 to 5, with 5 indicating a high degree of accuracy. The mean, as can be seen in Figure 3.12 was close to 4.5 for the entire sample across all observations. Fitting a polynomial trend line to the data has a better fit to the data and shows a slight decrease in accuracy towards the end of the study as would be expected due to fatigue as a result of extended wakefulness.

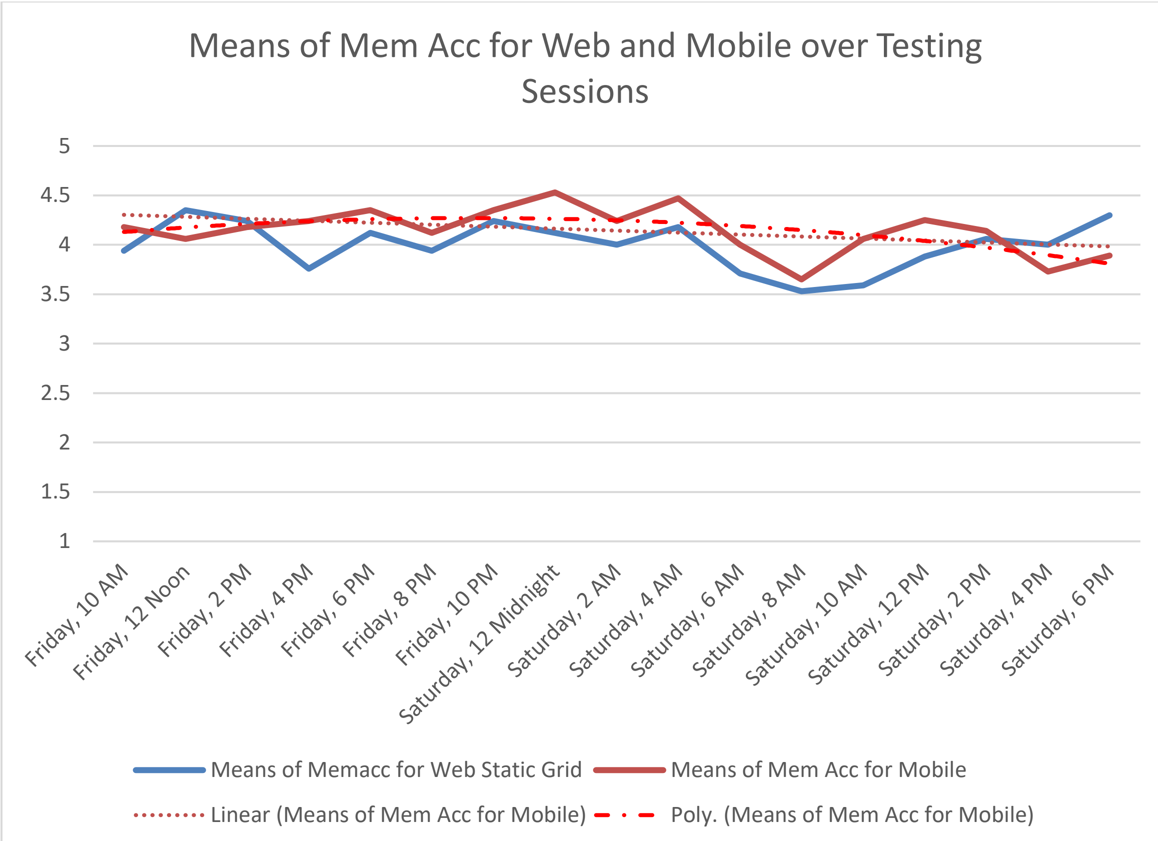

Figure 3.12 Memory Accuracy for Mobile Shape Recognition

In addition to examining the memory accuracy scores, we also looked at simple correlations between the AlertMeter® - Memory Accuracy scores (memacc) and the scores from the Working Memory Test Battery (WMTB) subtests at Time 1, or 10 a.m. on Friday, when the 36-hour sleep deprivation trial began. As can be seen in Figures 3.13 and 3.14 and Table 3.12, most of the significant positive correlations between AM – Memory Accuracy with the WMTB occurred at Time 1 at 10 a.m. and noon as would be expected if both the AlertMeter® memory items and the WMTB are assessing memory. This indicates that the AM Memory Accuracy items are working and provide a valid measure of memory. The Shape Pair Recognition items were not correlating as well with the initial WMTB.

Correlations of the Memory Accuracy and Shape Pair Recognition Accuracy with the Working Memory Test Battery (WMTB) were inconclusive and not consistent with expected patterns. Most correlations were not consistent with a decline in performance as wakefulness increased.

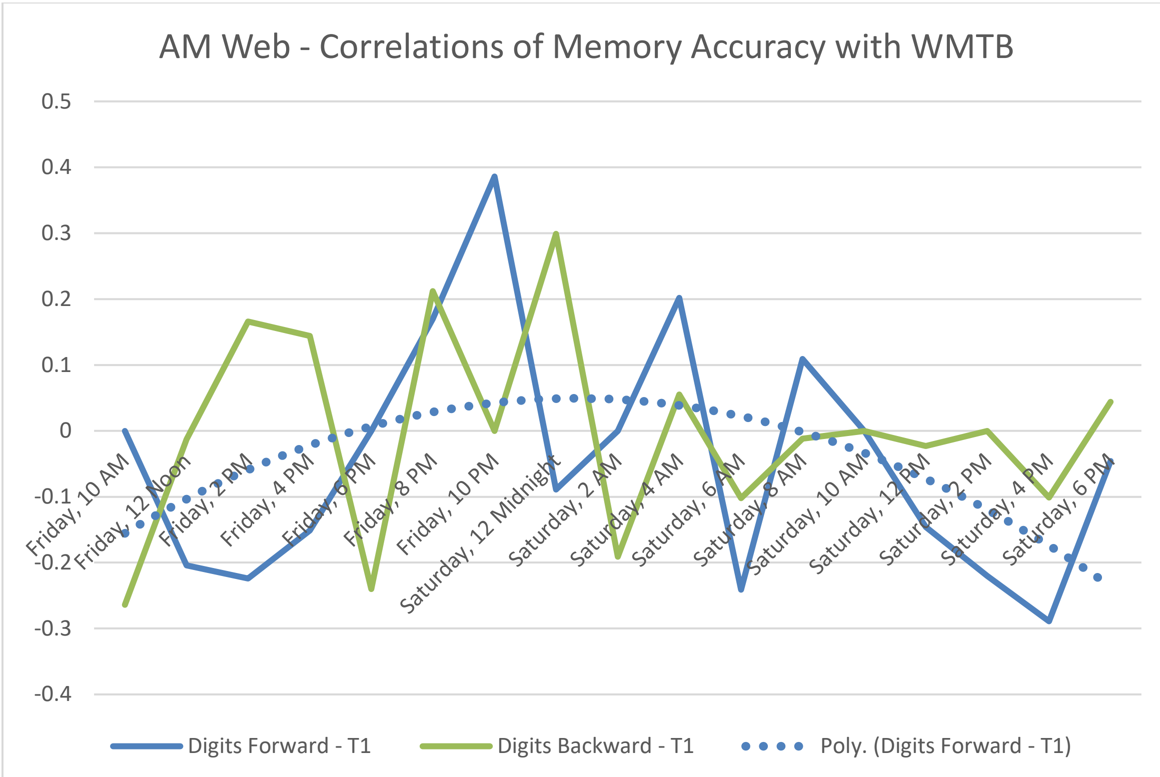

Figure 3.13 Memory Accuracy for AM Web/Tablet

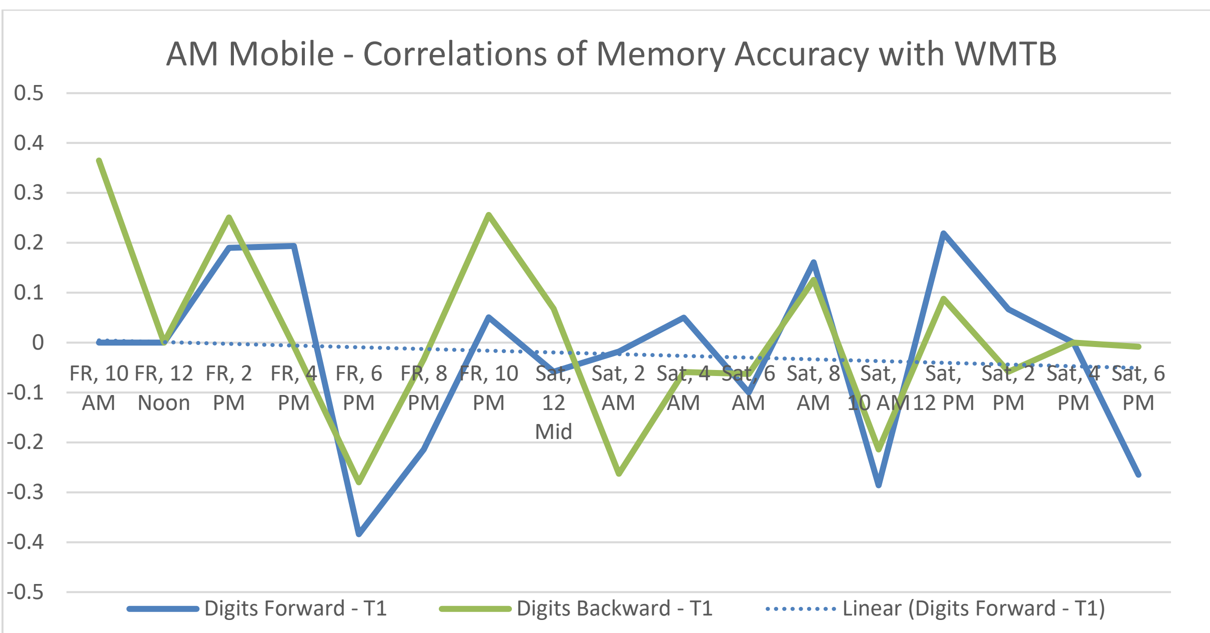

Figure 3.14 Memory Accuracy for AM Mobile

Previous research has shown that memory tasks require participants to encode and later retrieve stimuli, in contrast with vigilance tasks like the PVT (e.g., Lisper & Kjellberg, 1972; Yang, Lin, & Spielman, 2004), which require participants simply to pay attention and respond as quickly as possible to a stimulus change. In memory probe tasks indicating if a probe stimulus has occurred in a previously presented series of letters, words, or numbers are used. Memory tests have also involved a recognition test, although some researchers have used recall memory tests (Drake, Roehrs, Burduvali, Bonahoom, Rosekind, & Roth, 2001), spatial location memory tests (e.g., McEvoy, Smith, & Gevins, 2001, and working memory tests (Chee & Choo, 2004). Polzella (1975) found significant effects of sleep deprivation using a visual recognition task after a complete night of deprivation. Also, Wee, Asplund, and Chee (2017) found that visual short-term memory was decreased after 24 hours of sleep deprivation. They noted that sleep deprivation may inhibit the encoding of a visual image, but not the recall of a stored item. The AlertMeter® test items are based on a visual recognition process.

Overall, the AlertMeter® overall memory accuracy score appears to be demonstrating better validity with the WMTB than the shape pair recognition accuracy given that no clearly discernable pattern of correlations was obtained. It may be that the effects of practice and repetition are having an effect due to the finding of significance with both the Memory accuracy and the shape pair accuracy correlated with the time 1 baseline at 4 p.m. on Saturday. Additionally, the effects of boredom and lack of motivation might also have contributed to lack of significance at other testing times. However, further investigation of sensitivity and the nature of the stimuli recognized and encoded of the AlertMeter® tests may be necessary.

Table 3.12 Correlations AM Memory and Shape Recognition Accuracy and WMTB Subscales

AM Memory Accuracy with WMTB Time 1								
	DF	LDF	DB	LDB	DSS	LDSS	DSTSS	DSTot
FR 10 AM	.43*	0.41	0.37	0.36	0.40	0.30	.48*	.46*
FR 12 N	-0.49*	-0.49*	-0.42*	-0.43*	-0.11	-0.305	-0.42*	-0.41
FR 2 PM	0.19	0.21	0.25	0.11	.52*	.62**	0.40	0.35
FR 4 PM	0.19	-0.02	-0.01	0.03	0.18	0.20	0.12	0.15
FR 6 PM	-0.38	-0.35	-0.28	-0.33	-0.10	-0.17	-0.30	-0.34
FR 8 PM	-0.21	-0.28	-0.03	-0.06	-0.23	-0.28	-0.21	-0.17
FR 10 PM	0.05	-0.04	0.26	0.09	0.20	0.08	0.16	0.18
Sat 12 M	-0.06	0.08	0.07	-0.02	0.05	0.17	0.04	0.04
Sat 2 AM	-0.02	-0.10	-0.26	-0.25	-0.33	-0.27	-0.25	-0.24
Sat 4 AM	0.05	-0.07	-0.06	-0.12	-0.09	-0.07	0.02	-0.05
Sat 6 AM	-0.10	-0.05	-0.06	-0.22	-0.12	-0.08	-0.07	-0.12
Sat 8 AM	0.16	0.17	0.13	-0.02	0.41	.41*	0.33	0.24
Sat 10 AM	-0.29	-0.34	-0.21	-0.26	-0.24	-0.30	-0.18	-0.27
Sat 12 PM	0.22	0.12	0.09	0.28	0.40	0.42	0.23	0.25
Sat 2 PM	0.07	-0.04	-0.06	0.04	-0.06	-0.08	-0.22	0.00
Sat 4 PM	.57*	0.41	.66*	.66*	.78**	.71**	.77**	.78**
Sat 6 PM	-0.27	-0.15	-0.01	-0.07	-0.12	-0.27	-0.30	-0.24
AM Shape Pair Accuracy with WMTB Time 1								
FR 10 AM	0.14	-0.03	0.12	0.23	0.02	0.00	0.10	0.10
FR 12 N	-0.16	-0.33	-0.10	0.08	0.09	-0.05	0.02	-0.07
FR 2 PM	-0.10	-0.40	-0.06	0.02	0.12	-0.03	0.04	-0.03
FR 4 PM	-0.32	-0.36	-0.03	0.12	0.03	-0.17	-0.31	-0.13
FR 6 PM	-0.30	-0.30	-0.22	-0.06	-0.36	-0.38	-0.27	-0.33
FR 8 PM	0.02	-0.06	-0.10	-0.32	-0.02	0.00	-0.01	-0.02
FR 10 PM	0.39	0.21	0.22	0.20	0.27	0.25	0.34	0.32
Sat 12 M	-0.27	-0.19	-0.08	-0.17	0.01	0.01	-0.08	-0.10
Sat 2 AM	-0.18	-0.18	-0.46	-0.45	0.06	0.17	-0.15	-0.22
Sat 4 AM	-0.08	-0.25	-0.10	-0.16	0.07	-0.01	0.03	-0.01
Sat 6 AM	-0.12	-0.12	-0.33	-0.31	-.44*	-0.33	-0.27	-0.32
Sat 8 AM	-0.06	-0.18	-0.20	-0.12	0.11	0.14	-0.13	-0.06
Sat 10 AM	0.06	-0.05	-0.10	-0.11	-0.02	-0.08	-0.02	-0.03
Sat 12 PM	.44*	0.41	0.20	0.18	.46*	.61**	0.43	0.42
Sat 2 PM	-0.05	-0.06	-0.22	-0.16	-0.04	0.02	-0.23	-0.09
Sat 4 PM	0.50	0.41	0.47	.54*	.66*	.65*	.59*	.61*
Sat 6 PM	-0.11	-0.07	-0.09	0.05	0.51	0.38	0.07	0.10
* p<.05 ** p<.01 (1-tailed).								

DF – Digit Span Forward; LDF – Longest Digit Span Forward; DB – Digit Span Backward; LDB – Longest Digit Span Backward score; DSS – Digit Span Scaled Score; LDSS – Longest Digit Span Scaled Score; DSTSS – Digit Span Tota Score by age group; DSTot – Digit Span Total Score.

4. DISCUSSION

The results of the present study provide considerable support for the utility and validity of the AlertMeter® as a measure of alertness and fatigue. Measurement data collected from a sample of healthy adult male and female volunteers was found to show significant change in neuropsychological performance — reaction time — consistent with other published studies as wakefulness increased. These findings support the use of the AlertMeter® as a valid measure of the performance indicators associated with fatigue or alertness. Moreover, the results show that both the tablet-based device and the mobile (handheld iPhone or Android) based apps are highly similar in performance thus demonstrating the practical utility of the instruments.

4.1 AlertMeter® Correlation with the PVT

The results indicate a moderate degree of association between the PVT and the AlertMeter® and therefore evidence of concurrent validity for the measurement of fatigue and alertness in persons completing the AlertMeter® device. As with other previously published research (Lee, Bardwell, Israel, & Dimsdale, 2010; Grant, et. al., 2017; Price, Moore, Galway, & Linden, 2017) the PVT demonstrates a steady performance decrement over time as time awake increases. Similarly, the reaction time performance or latency on the AlertMeter® also demonstrates a sensitivity to measuring the cognitive deficits that are developed as fatigue and sleep deprivation increases. Because of differences in the different presentation format, and the cognitive neuropsychological processing of the more complex stimuli of the AM, it is not surprising that the two devices are not identical. The AM stimuli require more cognitive processing and thus a longer throughput to achieve a result. The PVT, considered the “gold standard” in the assessment of vigilance and reaction time in sleep deprivation studies, is very sensitive to the performance decrements associated with both acute and extended wakefulness. The AlertMeter® is also sensitive to these changes.

4.2 AlertMeter® Stimulus Items

The AM items, as reflected in the bin groupings, are randomly distributed for administration but include various levels of difficulty. The items are grouped into bins and the results indicate that not all the bins of items are equally sensitive in detecting fatigue. In some cases, there is a weak relationship between bins of items and the cumulative effect of sleep deprivation. For example, in Figure 8 and corresponding Table 7, Mobile AM reaction times and PVT reaction times, Bin 102, 201, 202 and 203 show little correspondence with PVT and PVT-Lapses but do show a significant correlation with KSS. Thus, while these bins of items are detecting the presence of subjective fatigue, they are less aligned with the simple reaction time measure of fatigue. Again, because of the order of administration of the bins, the confounding effects of practice and the utilization of different underlying cognitive neurological processes, the interpretation of the lack of correspondence with PVT is speculative. Nevertheless, additional refinement of the bin items and or the order of their administration might be worthy of further study. The overall results still stand, the AM is a valid measure of fatigue, however, it does not always function in the exact same fashion as the simple measure of reaction time, the PVT.

4.3 Memory – AlertMeter® Memory Correlations with WMBT Tests

As discussed at the outset, sleep deprivation is typically related to memory impairment (Karem, et. al., 2020). To assess the validity of the AM Memory tests, we correlated the Mobile Shape Recognition Memory Accuracy Scores with the WMTB. We collected data at three time points for the WMBT. Several subtests were combined to create a WMTB composite index. Using the T1, or 10 a.m., Friday morning as the baseline the AM Memory Accuracy and Shape Pair Accuracy scores were corelated with the Working Memory Test Battery (WMBT) scores. Unfortunately, the results of the correlations were inconclusive as the correlations and findings displayed no consistent interpretable pattern. These results suggest that the current AM Memory metrics are not functioning in the desired fashion in that they are not correlated with decreased memory accuracy over time. Several factors could be contributing to these findings. For example, the AM task of detecting and recalling the various shapes may not be sufficiently sensitive to fatigue and may not be deteriorating as expected. Some research suggests that if the task is more difficult, even sleep deprived study participants may momentarily increase their attention and concentration to ensure that they have obtained the correct answer (Lim & Dinges, 2010)[18].

Another factor that may account for inconsistent findings regarding the memory measure is that the AM is using a straight visual recognition and matching task. Components of the WMTB test are more related to short-term verbal memory. Visual memory is thought to be stored in the posterior visual cortex; whereas other short-term memory capabilities are associated with sensory-specific parietal and temporal areas of the brain (Muller & Knight, 2006).[19] However, these two types of working memory are consistently correlated with one another, and we would still expect a consistent pattern in the correlations between AM and WMTB. The AlertMeter® tasks involve identification of visual objects (symbols and shapes) and discrimination of the shapes based on their rotation and relative position. All together, these are complex tasks that may not easily correspond to the subscales of the WMTB. In fact, due to the unique nature of these tasks, they may present difficult challenges for any type of comparison and subsequent validation.

Another factor to consider relative to this issue is similarity of the AM task to real world occupational activity. The AM is primarily directed toward the cognitive requirements of driving or operating machinery. More often than not, work-related activity is concerned with letter and number recognition and decision making based on abstract concepts. Other occupational activities often involve arithmetical or mathematical concepts such as addition, multiplication etc. Therefore, some simplification of the AM tasks may lead to more immediate correspondence and sensitivity to degradation in cognitive performance as a function of sleep deprivation and fatigue. The AM is an effort, however imperfect, to compress as much testing into the smallest possible container and at the same time is not annoying to take. This limits test time to a maximum of two minutes and the annoyance factor should not create too much resistance from ordinary users.

[18] Lim, J., & Dinges, D. F. (2010). A meta-analysis of the impact of short-term sleep deprivation on cognitive variables. *Psychological Bulletin, 136*(3), 375–389. https://doi.org/10.1037/a0018883

[19] Müller, N., Knight, R. (2006) The functional neuroanatomy of working memory: Contributions of human brain lesion studies, Neuroscience, 139, 1, 51-58, ISSN 0306-4522, https://doi.org/10.1016/j.neuroscience.2005.09.018.

4.4 Study Limitations

Results of the present study may be limited by characteristics of the participating individuals. The study participants were predominantly highly educated college graduates. It has been reported that individuals with higher cognitive abilities may be less susceptible to memory impairment in sleep deprivation studies. In addition, some research has shown that individual differences in response to sleep deprivation can occur. Van Dongen, et. al. (2005) has written on this topic and proposed a research agenda. However, there may be a need to assess and identify persons with less sensitivity to fatigue to be selected for high-risk occupations. For example, the US Navy Seals are noted for selecting persons who can endure long periods of wakefulness. The presence of particular hardy individuals in the study participant population may have attenuated the robustness of the findings, which might have been more apparent in a less resilient group. Decreasing the variability of the RTs overall may have diluted the strength of the relationships.[20] The study was also conducted in a controlled environment, where research assistants assisted participants from napping or falling asleep. Accordingly, it is unknown how this environment will generalize to work settings. Yet, this design is the most optimal to demonstrate that the AlertMeter® is sensitive to change and associated with the most well-known measure of fatigue.

[20] Van Dongen, H.P., Vitellaro, K.M., & Dinges, D.F. (2005). Individual differences in adult human sleep and wakefulness: Leitmotif for a research agenda. Sleep. 2005 Apr;28(4):479-96.

5. CONCLUSIONS AND RECOMMENDATIONS

In conclusion, the AM was significantly correlated with the PVT Reaction time, PVT-Lapses, and subjective measures of alertness and fatigue from the Karolinska Sleepiness Scale all which showed decrements over time in association with increased amounts of wakefulness. Results of the study demonstrate concurrent validity of the AM with the PVT as a measure of fatigue and alertness. Moreover, reaction time magnitude varied as expected with increased amounts of wakefulness and in accordance with expected circadian patterns, further supporting the validity of AlertMeter® as a measure of alertness and fatigue. Findings were inconclusive with respect to the memory assessment as no consistent interpretable results were obtained that demonstrated a correlation with the WMTB standard measure of memory and recall.

6. RECOMMENDATIONS

1. Additional research using a broader range of representatives from transportation crafts and modes would be of benefit in establishing the basis for acceptance of this methodology in the field. Transportation employees are somewhat leery of instrumentation and techniques that provide detailed information on their physical and mental well-being to their employers.

2. Further study of the usefulness of this methodology with persons operating equipment while adhering to a variable start time schedule would also be beneficial. For the most part, a significant number of railroad operations personnel work a schedule where they are continually on call with no set start times. These are usually called “road” or “pool” jobs that are physically and mentally demanding. Additional real time in the field data would be useful in establishing the utility and efficacy of this technology.

3. A demonstration of this technology within a labor organization would also be useful in establishing its credibility and efficacy. Once a large number of employees or members of a labor organization have adopted the approach, there will be wider acceptance.

4. Work to increase the robustness of the assessment methodology would be useful in creating a set of stimulus items (bins) that are most sensitive to fatigue. It may be possible to rearrange the stimulus items and their order of administration to obtain an even more sensitive assessment measure.

REFERENCES

1. Alhola, P., & Polo-Kantola, P. (2007). Sleep deprivation: Impact on cognitive performance. *Neuropsychiatric disease and treatment*, *3*(5), 553–567.
2. 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 & Prevention*, Volume 126, 2019, Pages 173-176, ISSN 0001-4575, https://doi.org/10.1016/j.aap.2017.11.041.
3. Arsintescu, L., Mulligan, J. B., & Flynn-Evans, E. E. (2017). Evaluation of a Psychomotor Vigilance Task for Touch Screen Devices. *Human Factors*, 59(4), 661–670. https://doi.org/10.1177/0018720816688394
4. Balkin TJ, Bliese PD, Belenky G, Sing H, Thorne DR, Thomas M, Redmond DP, Russo M, Wesensten NJ. (2004) Comparative utility of instruments for monitoring sleepiness-related performance decrements in the operational environment. *J Sleep Res*. 2004 Sep;13(3):219-27. doi: 10.1111/j.1365-2869.2004.00407.x. PMID: 15339257.
5. Bioulac, S., Micoulaud-Franchi, J. , Arnaud, M., Sagaspe, P., Moore, N., Salvo, F., Philip, P. (2017). Risk of Motor Vehicle Accidents Related to Sleepiness at the Wheel: A Systematic Review and Meta-Analysis, *Sleep*, Volume 40, Issue 10, October 2017, zsx134, https://doi-org.du.idm.oclc.org/10.1093/sleep/zsx134
6. Brunet, JF., Dagenais, D., Therrien, M., Gartenberg, D., and Forest, G. (2017) Validation of sleep-2-peak: A smartphone application that can detect fatigue related changes in reaction times during sleep deprivation. *Behavioral Research*. (2017) 49: 1460. https://doi.org/10.3758/s13428-016-0802-5
7. Chee, M., & Choo, W. C. (2004) Functional imaging of working memory after 24 hr of total sleep deprivation. *The Journal of Neuroscience*, 24, 4560-4567.
8. Dawson, D. (2014) Look before you sleep: Evaluating the use of fatigue detection technologies within a fatigue risk management system for the road transport industry. *Sleep Medicine Reviews* 18, 2, 2014, 141-152.
9. Dawson, Drew, Darwent, David, and Roach, Gregory D. (2017) "How Should a Bio-mathematical Model Be Used within a Fatigue Risk Management System to Determine Whether or Not a Working Time Arrangement Is Safe?" *Accident Analysis and Prevention* 99.Pt B (2017): 469-73. Web.
10. Donovan KA, & Jacobsen PB. (2010). The Fatigue Symptom Inventory: a systematic review of its psychometric properties. *Support Care Cance*r. 2010 Feb;19(2):169-85. DOI: 10.1007/s00520-010-0989-4. Epub 2010 Sep 8.
11. Donovan KA, Stein KD, Lee M, Leach CR, Ilozumba O, Jacobsen PB. (2015). Systematic review of the multidimensional fatigue symptom inventory-short form. *Support Care Cancer*. 2015 Jan;23(1):191-212. DOI: 10.1007/s00520-014-2389-7. Epub 2014 Aug 22.
12. DOT Strategic Plan. https://www.transportation.gov/dot-strategic-plan (Aug 16, 2018.)
13. Drake, C., Roehrs, T., Burduvall, E., Bonahoom, A., Rosekind, M., & Roth, T. (2001) Effects of rapid versus slow accumulation of eight hours of sleep loss. *Psychophysiology*, 38, 979-987.

14. Flynn-Evans, E.E., Kirkley, C., Young, M. et al. (2020) Changes in performance and bio-mathematical model performance predictions during 45 days of sleep restriction in a simulated space mission. *Sci Rep* 10, 15594 (2020). https://doi.org/10.1038/s41598-020-71929-4
15. Frenda, Steven J, and Fenn, Kimberly M. "Sleep Less, Think Worse: The Effect of Sleep Deprivation on Working Memory." Journal of Applied Research in Memory and Cognition 5.4 (2016): 463-69. Web.
16. Glenville, M., Broughton. R.,Wing, A., & Wilkinson, R. (1978) Effects of sleep deprivation on short duration performance measures compared to the Wilkinson auditory vigilance task. *Sleep*, 1, 169-176.
17. Golz, M., Sommer, D., Trutschel, U. et al. Evaluation of fatigue monitoring devices. *Somnology* (2010) 14: 187. Https://doi.org/10.1007/s11818-010-0482-9.
18. Gottlieb, D. J., Ellenbogen, J. M., Bianchi, M. T., & Czeisler, C. A. (2018). Sleep deficiency and motor vehicle crash risk in the general population: a prospective cohort study. *BMC medicine*, *16*(1), 44. doi:10.1186/s12916-018-1025-7.
19. Grant, D.A., Honn, K.A., Layton, M.E., Riedy, S. M., and Van Dongen, H. P.A. (2017). A 3-minute smartphone-based and tablet-based psychomotor vigilance tests for the assessment of reduced alertness due to sleep deprivation. *Behav Res* (2017) 49: 1020. https://doi.org/10.3758/s13428-016-0763-8
20. https://www.scpr.org/news/2017/11/13/77705/fatigue-poor-safety-caused-deadly-california-bus-c/
21. Hoddes, E., Zarcone, V., & Dement, W. (1972). Stanford sleepiness scale. *Encyclopedia of Sleep Medicine* , *1184* .
22. Jongstra S, Wijsman LW, Cachucho R, Hoevenaar-Blom MP, Mooijaart SP, Richard E (2017). Cognitive Testing in People at Increased Risk of Dementia Using a Smartphone App: The iVitality Proof-of-Principle Study. JMIR Mhealth Uhealth 2017;5(5):e68 URL: https://mhealth.jmir.org/2017/5/e68 DOI: 10.2196/mhealth.6939 PMID: 28546139 PMCID: 5465383
23. Kaida, K., Takahashi, M., Åkerstedt, T., Nakata, A., Otsuka, Y., Haratani, T., & Fukasawa, K. (2006). Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clinical Neurophysiology*, 117, 1574–1581.
24. Lee, I.-S., Bardwell, W. A., Ancoli-Israel, S., & Dimsdale, J. E. (2010). Number of Lapses during the Psychomotor Vigilance Task as an Objective Measure of Fatigue. *Journal of Clinical Sleep Medicine* , 6(2), 163–168.
25. Lifesavers National Conference on Highway Safety Priorities, Mark R. Rosekind, Ph.D., Administrator, NHTSA (March 16, 2015).
26. Lim, J. & Dinges, D.F. (2008) Sleep Deprivation and vigilant attention. Annals of the New York Academy of Sciences, 1129: 305-322 doi:10/1196/annals.1417.002
27. Lim, J., & Dinges, D. F. (2010). A meta-analysis of the impact of short-term sleep deprivation on cognitive variables. *Psychological Bulletin*, 136(3), 375–389. https://doi-org.du.idm.oclc.org/10.1037/a0018883.supp (Supplemental)

28. Lisper, H., & Kellberg, A. (1972) Effects of 24-hour sleep deprivation on rate of decrement in a 10-minute auditory reaction time task. *Journal of Experimental Psychology*, 96, 287-290.
29. Maslovat, D., Klapp, S. T., Jagacinski, R. J., & Franks, I. M. (2014). Control of response timing occurs during the simple reaction time interval but on-line for choice reaction time. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(5), 2005–2021. https://doi-org.du.idm.oclc.org/10.1037/a0037522
30. McEvoy, L., Smith, M., & Gevins, A. (2000) Test-retest reliability of cognitive EEG. *Clinical Neurophysiology*, 1 1 1, 457-463.
31. Müller, N., Knight, R. (2006) The functional neuroanatomy of working memory: Contributions of human brain lesion studies, *Neuroscience*, 139, 1, 51-58, ISSN 0306-4522, https://doi.org/10.1016/j.neuroscience.2005.09.018
32. National Center for Statistics and Analysis. (2011, March). Traffic Safety Fact Crash*Stats: Drowsy Driving. (DOT HS 811 449). Washington, DC: NHTSA. Retrieved from www-nrd.nhtsa.dot.gov/pubs/811449.pdf
33. NTSB. (2017) Motor coach Collision with Combination Vehicle Palm Springs, California https://www.ntsb.gov/news/events/Documents/2017-HWY17MH005-BMG-abstract.pdf
34. NTSB. (2018). Long Island Rail Road Passenger Train Strikes Platform in Atlantic Terminal. https://ntsb.gov/investigations/AccidentReports/Reports/RAB1802.pdf
35. NTSB. Federal Regulator Not Doing Enough to Keep Unsafe Operators Off Roads. https://www.ntsb.gov/news/press-releases/Pages/PR20171113.aspx
36. Polzella, D.(1975) Effects of sleep deprivation on short-term recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, 104, 194-200.
37. Price E, Moore G, Galway L, Linden M. (2017) Validation of a Smartphone-Based Approach to In Situ Cognitive Fatigue Assessment. *JMIR Mhealth Uhealth* 2017;5(8):e125. URL: https://mhealth.jmir.org/2017/8/e125 DOI: 10.2196/mhealth.6333 PMID: 28818818 PMCID: 5579321.
38. Raidy, Desiree J, and Scharff, Lauren F V. "Effects of Sleep Deprivation on Auditory and Visual Memory Tasks." *Perceptual and Motor Skills* 101.2 (2005): 451-67. Web. https://doi-org.du.idm.oclc.org/10.2466/pms.101.2.451-467
39. Roach, Gregory D, Roberts, Paul, Dawson, Drew, Ferguson, Sally, Meuleners, Lynn, Brook, Libby, and Sargent, Charli. (2017). "Controlling Fatigue Risk in Safety-critical Workplaces: A Summary of Selected Papers from the 9th International Conference on Managing Fatigue in Transportation, Resources and Health." *Accident Analysis and Prevention,* 99.Pt B (2017): 379-82. Web.
40. Santisteban, Jose Arturo, Brown, Thomas G, Ouimet, Marie Claude, and Gruber, Reut. "Cumulative Mild Partial Sleep Deprivation Negatively Impacts Working Memory Capacity but Not Sustained Attention, Response Inhibition, or Decision Making: A Randomized Controlled Trial." Sleep Health 5.1 (2019): 101-08. Web. https://doi.org/10.1016/j.sleh.2018.09.007
41. Shuyan, M. & Gangtie, L. (2009). *Expert Systems with Applications*, May 2009, Vol.36(4), pp.7651-7658.

42. St Hilaire, Melissa A, Rüger, Melanie, Fratelli, Federico, Hull, Joseph T, Phillips, Andrew J K, and Lockley, Steven W. "Modeling Neurocognitive Decline and Recovery During Repeated Cycles of Extended Sleep and Chronic Sleep Deficiency." *Sleep*, 2017-01-01, Vol.40 (1). Web.
43. Tefft, B. C. (2012). Prevalence of motor vehicle crashes involving drowsy drivers, United States, 1999-2008. *Accident Analysis & Prevention*, 45(1): 180-186.
44. Van Dongen, H. P. A., Maislin, G., Mullington, J. M., & Dinges, D. F. (2003). The cumulative cost of additional wakefulness: Dose–response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep*, 26, 117–129.
45. Van Dongen, H.P., Vitellaro, K.M., & Dinges, D.F. (2005). Individual differences in adult human sleep and wakefulness: Leitmotif for a research agenda. *Sleep.* 2005 Apr;28(4):479-96.
46. Wang,X., & Xu, C. (2016) Driver drowsiness detection based on non-intrusive metrics considering individual specifics. *Accident Analysis & Prevention*, Volume 95, Part B, 2016, pp. 350-357
47. Wee, Natalie, Asplund, Christopher L, and Chee, Michael W L. "Sleep Deprivation Accelerates Delay-related Loss of Visual Short-term Memories without Affecting Precision." *Sleep,* 36.6 (2013): 849-56. Web. doi.org/10.5665/sleep.2710
48. Wilkinson VE; Jackson ML; Westlake J; Stevens B; Barnes M; Swann P; Rajaratnam SMW; Howard ME. (2013) The accuracy of eyelid movement parameters for drowsiness detection. *Journal of Clinical Sleep Medicine*, 2013;9(12):1315-1324.
49. Yang, C. M., Lin, F. W., & Spielman, A. J. (2004) A standard procedure enhances the correlation between subjective and objective measures of sleepiness. *Sleep*, 27, 329-332.