

Sleep and Performance Measures in Soldiers Undergoing Military Relevant Training

Approved for public release, distribution unlimited.

**U.S. Army
Aeromedical Research
Laboratory**

US A A R L

Notice

Qualified requesters

Qualified requesters may obtain copies from the Defense Technical Information Center (DTIC), Cameron Station, Alexandria, Virginia 22314. Orders will be expedited if placed through the librarian or other person designated to request documents from DTIC.

Change of address

Organizations receiving reports from the U.S. Army Aeromedical Research Laboratory on automatic mailing lists should confirm correct address when corresponding about laboratory reports.

Disposition

Destroy this document when it is no longer needed. Do not return it to the originator.

Disclaimer

The views, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other official documentation. Citation of trade names in this report does not constitute an official Department of the Army endorsement or approval of the use of such commercial items.

Human use

Human subjects participated in these studies after giving their free and informed voluntary consent. Investigators adhered to AR 70-25 and USAMRMC Reg 70-25 on Use of Volunteers in Research.

REPORT DOCUMENTATION PAGE					<i>Form Approved OMB No. 0704-0188</i>	
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.						
1. REPORT DATE (DD-MM-YYYY) 26-06-2009		2. REPORT TYPE Final			3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE Sleep and Performance Measures in Soldiers Undergoing Military Relevant Training					5a. CONTRACT NUMBER	
					5b. GRANT NUMBER	
					5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) William D. Killgore (WRAIR) Arthur Estrada (USAARL) Tiffany Rouse (USAARL) Robert M. Wildzunas (USAARL) Thomas J. Balkin (WRAIR)					5d. PROJECT NUMBER	
					5e. TASK NUMBER	
					5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Aeromedical Research Laboratory P.O. Box 620577 Fort Rucker, AL 36362-0577					8. PERFORMING ORGANIZATION REPORT NUMBER USAARL 2009-13	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command 504 Scott Street Fort Detrick, MD 21702					10. SPONSOR/MONITOR'S ACRONYM(S) USAMRMC	
					11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited.						
13. SUPPLEMENTARY NOTES						
14. ABSTRACT Inadequate sleep is known to impair a variety of cognitive capacities, including attention, vigilance, concentration, and aspects of higher order reasoning and judgment. The ability to unobtrusively measure fatigue and predict its effects on cognitive performance is vital to successful military operations. Wrist actigraphy's ability to accurately measure and predict performance in militarily relevant activities is not well validated. Healthy military volunteers (N = 108) wore wrist activity monitors while undergoing military training. Actigraphic data were analyzed and used to predict academic success. Regardless of course type or test content, academic performances were significantly predicted by total sleep time, sleep latency, number of immobile minutes, and sleep fragmentation, but not total activity. Academic performance was significantly related to the amount and quality of sleep obtained within the 48-hour period preceding the exams. Actigraphy appears to be a valid and unobtrusive method for predicting academic performance in military courses, although participant compliance and detection of off-wrist periods need to be improved.						
15. SUBJECT TERMS actigraphy, actiwatches, performance prediction						
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES 56	19a. NAME OF RESPONSIBLE PERSON Loraine Parish St. Onge, PhD	
a. REPORT UNCLAS	b. ABSTRACT UNCLAS	c. THIS PAGE UNCLAS			19b. TELEPHONE NUMBER (Include area code) 334-255-6906	

Reset

Acknowledgements

The authors would like to express their sincere gratitude to the following people for their contributions to this project.

- Ms. Elizabeth Stokes for help with all matters administrative.
- Dr. Loraine St. Onge and Dr. John S. Crowley for their editorial assistance.

Table of contents

	<u>Page</u>
Introduction.....	1
Background.....	1
Research objectives.....	4
Methods	5
General.....	5
Study population and description	5
Measures	6
Procedure	7
Data preparation and analysis	8
Data retrieval.....	8
Initial data cleaning.....	8
Overall analysis plans	8
Actiware analysis.....	10
Overview.....	10
Data extraction and sleep analysis	10
Data quality check	12
Visual quality inspection of the data.....	12
Automated quality inspection of the data	12
Final sample after data cleaning	14
Output	14
SHARP analysis.....	14

Table of contents (continued)

	<u>Page</u>
Procedure	14
SHARP output	15
Smoothing iterations	16
Results	18
Statistical analyses	18
Actiware analyses	18
Average total sleep per night	18
Comparisons across class groups.....	19
Total sleep time.....	19
Sleep efficiency.....	20
Sleep latency	21
Sleep fragmentation	22
Prediction of exam performance from actigraphy	23
Overview.....	23
Total sleep prior to Exam 1.....	23
Relationship between total sleep and performance.....	25
Relationship between sleep latency and performance	26
Relationship between other actigraphic indices and performance.....	27
Assumed sleep	27
Actual wake time	27
Sleep efficiency.....	27
Number of sleep bouts	27

Table of contents (continued)

	<u>Page</u>
Immobile minutes	27
Fragmentation index	28
Validation of SHARP	28
Overview.....	28
Prediction of exam scores from SHARP ME index	28
Comparison of exam scores among SHARP ME color zones.....	29
Prediction of exam scores from SHARP "Percent Below Criterion" index	29
Prediction of exam scores from SHARP Typical Effectiveness (TE) index	29
SHARP Percent Change in Effectiveness (PCE) scores.....	31
Prediction of exam scores from SHARP Average Sleep index	33
All three exams combined	34
Prediction of all three exams scores from SHARP ME index	34
Prediction of all three exams scores from SHARP Average Sleep index	35
SHARP PCE scores for all three exams	36
Discussion.....	37
Limitations	40
Conclusions.....	40
References.....	41
Manufacturers List.....	44

Table of contents (continued)
List of figures

	<u>Page</u>
1. General study design.....	7
2. Using average measured sleep during the 2 days preceding each examination to predict exam performances.....	9
3. Using average measured sleep during the 1 day preceding each examination to predict exam performances.....	9
4. Using the SHARP data analysis to examine the ability of the sleep-performance prediction model for all days preceding each examination to predict exam performances	9
5. Actiware 3.41 Sleep Analysis window	11
6. Example of the daily sleep scores extracted for one subject	11
7. Examples of valid, probably invalid, and definitely invalid actigrams by visual inspection ..	12
8. SHARP graphical user interface	15
9. Example of the two smoothing levels on the data	17
10. Histogram of the mean hours of sleep obtained by all subjects for the entire duration of the study	18
11. Mean total sleep durations for each course group (significant differences exceeding $p < .05$ for LSD exam are indicated by red brackets)	19
12. Mean sleep efficiency (in percent) for each course group (significant differences exceeding $p < .05$ for LSD exam are indicated by red brackets)	20
13. Mean sleep latency for each course group (significant differences exceeding $p < .05$ for LSD exam are indicated by red brackets)	21
14. Mean sleep fragmentation index for each course group (significant differences exceeding $p < .05$ for LSD exam are indicated by red brackets)	22
15. Histograms showing the amounts of sleep obtained 2 days and 1 day prior to Exam 1.....	24
16. Scatterplots showing the significant positive correlation between exam scores and the amount of sleep obtained 2 days and 1 day prior to Exam 1	25

Table of contents (continued)
List of figures (continued)

	<u>Page</u>
17. Scatterplots showing the significant negative correlation between exam scores and the latency to fall asleep during the 2 days prior to Exam 1.....	26
18. Scatterplots showing the significant positive correlations between Exam 1 scores and the SHARP ME score at the three smoothing levels	29
19. Scatterplots showing the correlations between Exam 1 scores and the SHARP Percent Below Criterion (ME 90) at the three smoothing levels	30
20. Scatterplots showing the significant positive correlations between Exam 1 scores and the SHARP TE scores (three smoothing levels).....	31
21. Scatterplots showing the significant positive correlations between Exam 1 scores and the SHARP PCE scores (three smoothing levels)	32
22. Exam scores on Exam 1 differed for the three groups defined on the magnitude of decline in PCE scores	33
23. Scatterplots showing the significant positive correlations between Exam (Test) 1 scores and the SHARP Average Sleep scores prior to the exam (three smoothing levels)	34
24. Scatterplots showing the significant positive correlations between all three exam scores and the SHARP ME score (three smoothing levels).....	35
25. Scatterplots showing the significant positive correlations between all three exam scores and the SHARP Average Sleep scores prior to each exam (three smoothing levels).....	36
26. Color-coded categories defined on the magnitude of decline in Percent Change in Effectiveness scores across all three examinations combined.....	37

List of tables

1. Recruited participants by course and MOS	6
2. Kappa coefficients	13
3. SHARP output table.....	16
4. Group statistics for total sleep time in hours	19
5. Group statistics for sleep efficiency percent.....	20

Table of contents (continued)
List of tables (continued)

	<u>Page</u>
6. Group statistics for sleep latency in hours	21
7. Group statistics for sleep fragmentation index	22

Introduction

The quantity of rest and sleep obtained by ground and air warfighters are primary factors that determine their levels of alertness and performance. Providing information regarding the performance potential/status of individual Soldiers and groups of Soldiers would enhance military leaders' abilities to select the most suitable Soldier(s) for a given task/mission in order to maximize mission effectiveness. This research facilitates modeling for the Walter Reed Army Institute of Research (WRAIR) Fatigue Intervention and Recovery Model (FIRM) and the U.S. Army Research Institute of Environmental Medicine (USARIEM) Warfighter Physiological Status Monitoring (WPSM) programs and validates models for the Soldier Health and Performance Status Army Technology Objective (ATO). Validation of performance prediction models could assist military commanders and decision-makers in choosing proper courses of action during the planning, training, and execution phases of military operations.

Around-the-clock military operations today are the norm rather than the exception (Miller, 2005); with night operations a significant component of combat and training missions (Comperatore et al., 1993). Soldiers are often required to work for long periods of time without rest. This lack of rest can degrade Soldiers' ability to perform their duties efficiently, correctly, and in some cases, safely (Caldwell and Caldwell, 1993). In addition, Lieberman et al. (2005) report that there is considerable anecdotal documentation that combat-like stress can have a deleterious impact on the ability of warfighters to process cognitive information and act quickly, effectively, and decisively on the battlefield.

Background

Assessing performance under combat-like activity and stress is critical for military field studies. Activity monitors are one of the few devices that are capable of monitoring activity and behavior in free-living humans in the field on a minute-by-minute basis. These non-invasive monitors, principally worn on the wrist, have been used to produce graphical representations of subject activity (actigraphy) and have been successfully employed in Army field studies to assess sleep and performance (Lieberman and Coffey, 2000). During one such study in which the efficacy of melatonin was tested, Comperatore et al. (1996) were able to document Soldier adaptation to new work schedules by correlating actigraphy and Soldiers' cognitive performance.

Researchers have demonstrated the ability of actigraphs to predict sleep versus wakefulness for many years (Kripke et al., 1981; Webster et al., 1982; Cole and Kripke, 1988; Sadeh et al., 1989). Caldwell and Caldwell (1993) cautioned that although wrist activity monitors (WAMs) are good instruments for estimating sleep time, they cannot be used to determine the quality of sleep. Their study showed that when compared with the electroencephalograph (EEG) the WAM agreed 89 percent of the time. Error in the WAM tended to overestimate sleep by missing some brief arousals, one measure of sleep quality. According to Carskadon and Dement (1989), the WAM may be able to estimate slow wave sleep (SWS) better than the other stages since this stage is quieter, with little or no movement. This being said, Caldwell and Caldwell recommended that whenever actigraphs are used, "a conservative interpretation of the data is to discuss *rest time* instead of *sleep time*" and concluded that the WAM appeared to be a good tool for estimating the amount of sleep a person receives in situations where an EEG is not

practicable. Hence, since EEGs are not practical for field studies, WAMs are considered and accepted by the scientific community as reasonable substitutes.

One such WAM is the Actiwatch[®], produced by Mini Mitter, a Respironics Company, and is the data collection device used in this effort. According to the Actiwatch[®] Instructional Manual (2001), the Actiwatch[®] is a small, lightweight, limb-worn, device which utilizes an accelerometer to monitor the occurrence and degree of motion. The sensor integrates the degree and speed of motion and produces an electrical current that varies in proportional magnitude at a sampling rate of up to 32 Hz. It contains an omnidirectional sensor and is thus, sensitive to motion in all directions. Once collected, the data is wirelessly downloaded to a reader which is connected to a personal computer. Accompanying Actiwatch[®] software allows the manipulation, analysis, and presentation of the data. It has been designed with a water-resistant case for use at pressures up to 1 atmosphere.

Today's warfighters may be required to operate technically sophisticated modern weaponry under conditions of sleep deprivation, sleep inertia (the state of grogginess and disorientation commonly experienced on waking from sleep), and the "fog of war." All of these conditions have detrimental effects on human performance and cognition (Wertz et al. 2006; Lieberman et al., 2005; Belenky et al., 2003; LeDuc, 2000; Caldwell et al., 1999). Kruger (1989, summarized by Caldwell et al.) reviewed numerous studies on the effects of sustained work and sleep loss, and indicated that sleep deprivation: 1) increases mental lapses which have an impact on the speed and accuracy of responses; 2) reduces ability to acquire and recall information in complex tasks; 3) produces changes in brain activity associated with decreased alertness; and 4) slows cognitive ability in which task performance declines in conjunction with mood and motivation.

Models have been developed which promise to provide military leaders and individual warfighters with performance potential and/or status. Two such models are WRAIR's Sleep Performance Model (SPM), a mathematical algorithm used to predict performance based on prior sleep and circadian rhythm (Balkin et al., 2000) and another well-validated and extensively studied model which has been developed by investigators at WRAIR in collaboration with Scientific Applications International Corporation (SAIC[®]), known as the Sleep, Activity, Fatigue, and Task Effectiveness (SAFTE[™]) Model (Hursh et al., 2004).

In short, the SAFTE[™] model predicts a Soldier's "performance effectiveness" as a function of several factors, including the amount of time spent awake, the amount, quality, and timing of sleep, circadian rhythm influence (i.e., time of day effects), and sleep inertia (i.e., temporary grogginess after awakening). The SAFTE[™] model proposes that every individual has a "sleep reservoir" for cognitive performance that is depleted by continuous wakefulness and restored by sleep. When fully rested, the reservoir is full, but slowly empties at a rate of about 1 percent for every hour the person remains awake. Sleep refills the reservoir, but does so in a non-linear manner. The recuperative value of each minute of sleep (i.e., how fast the sleep reservoir refills) depends on how much remains in the reservoir at the time of sleep onset. The more depleted the reservoir, the more rapidly that each minute of obtained sleep will restore performance. Specifically, recuperation occurs at a faster rate in the early part of the sleep period and provides progressively less recuperative value as the sleep period continues. Similarly, the longer a person has gone without sleep (i.e., the more depleted the reservoir), the more rapidly the

reservoir will be filled during the early hours of the sleep period. The closer the reservoir is to its capacity, the greater the predicted performance effectiveness of the individual. Performance effectiveness, however, is also impacted by sleep inertia, or the temporary sluggishness in cognitive ability that occurs immediately upon awakening and which may persist for some time thereafter. The SAFTE™ Model predicts that sleep inertia produces the greatest degree of performance impairment immediately upon awakening and dissipates exponentially over the next two hours. Finally, predicted effectiveness from the SAFTE™ Model also depends on the time of day, or circadian rhythm. Performance normally follows a modest sinusoidal pattern throughout the day, with peaks in the late morning and again in the early evening. The SAFTE™ Model mathematically combines all of these components to provide a prediction of “cognitive performance effectiveness” for any time of day, expressed as a percentage of fully rested cognitive capacity. This index represents the fundamental capacity to perform effectively on tasks requiring perceptual discrimination, rapid response speed, efficient information processing, reasoning, and linguistic processing (Hursh et al., 2004).

The SAFTE™ Model has recently been implemented within a computerized tool to help leaders, decision-makers, and workers plan sleep/wake schedules more effectively. The tool allows the user to easily enter a hypothetical sleep/wake schedule or actual sleep/wake data from actigraphy and quickly create an individualized cognitive performance output. This program, the Fatigue Avoidance Scheduling Tool (FAST™), provides a user friendly interface and presents effectiveness scores in an intuitive graphical format. The FAST™ is available by contacting the developers at www.FASTSleeper.com. In the FAST™ program, predicted performance effectiveness for each individual is graphed as a continuous line that depicts minute by minute the expected cognitive performance effectiveness score according to the model parameters discussed previously. Scores are easily interpreted as falling within color coded ranges (red, yellow, green) representing various performance criteria. The program also includes a feature that “flags” significant factors that may pose fatigue-related safety concerns (e.g., time of day effects, cumulative sleep debt). The FAST™ program is particularly useful for studying the sleep/wake schedule of an individual over several days or weeks to determine how schedule shifts, chronic sleep restriction, time zone changes, and other events affect the level of performance effectiveness.

While the level of detail and analysis provided for any particular individual is one of its most attractive features of FAST™, this degree of detail can also make it unwieldy when one is interested in summarizing the performance of multiple individuals simultaneously. The FAST™ requires navigation through several screens and button clicks in order to import actigraph data and run an analysis on a single individual. Further, in its present form, FAST™ presents the output for each individual on a separate screen, again making it difficult to compare the performance of more than a few individuals at a time. Thus, while FAST™ can currently provide a tool for intensively studying the relationship between sleep schedules and predicted cognitive performance on an individual basis, making it an ideal tool for sleep researchers and health care professionals, it may not always be practical for use in a field expedient environment such as combat. Under operational conditions, military leaders often need a rapid and unambiguous summary of the current and predicted status of an entire unit or cohort of troops.

To address the difficulties of implementing the FAST in a field combat environment, Dr. William Killgore and colleagues at WRAIR have developed a program known as the Sleep History and Readiness Predictor (SHARP), which combines individual actigraphic sleep data to produce group (military unit) data. The SHARP provides the military commander with an analysis of the unit's mission effectiveness potential in the context of time of day. In other words, the commander is presented with a color-coded (red-yellow-green) unit summary and graph which targets a specific mission window during which his Soldier(s) is/are optimally prepared. Note that the WRAIR SPM is being integrated into a Sleep Watch. The Sleep Watch, under development at the WRAIR, is a "fuel gauge" that displays a performance prediction ranging from 0 to 100 percent of well-rested levels and a simple red-yellow-green analog scale that can alert the wearer of the need to acquire additional sleep.

The above described predictive models were developed using Soldier performance data obtained from the Psychomotor Vigilance Task (PVT). The PVT is a portable simple reaction time exam and is known to be sensitive to the effects of fatigue and sleepiness (Dinges et al., 1997). It displays a 3-millimeter (mm) light in a window for up to 1.5 seconds (s) during which time the subject responds by pressing a microswitch which records reaction time to the stimulus. The interstimulus interval varies randomly from 1 to 10 s. Balkin et al. (2000) write that the PVT was deemed optimal for the purpose of modeling the SPM because:

"1) there were no apparent learning effects; 2) the measure was sensitive to the experimental manipulation (i.e., there was adequate separation in mean performance levels between the various sleep groups); and 3) although fatigue might affect PVT performance (and account for some of its sensitivity to sleep loss), it is a short-duration (10 minutes) – thus, fatigue would be expected to account for a relatively small portion of the variance."

Hence, the SPM parameters were optimized using PVT data. The purpose of this study was to serve as a validation for the models developed from PVT data. This study's data, collected during operational military training conditions, is used to confirm the models' predictive qualities against operationally relevant performance scores.

Research objectives

The objectives of this study were to:

1. demonstrate that actigraphy can be obtained reliably and unobtrusively from Soldiers in militarily relevant settings (non-laboratory conditions, including strenuous physical activity);
2. determine the effectiveness of actigraphy as a predictor of cognitive performance under non-laboratory conditions (confirming PVT-based laboratory prediction models);
3. validate the use of SHARP for providing useful summary scores regarding sleep/wake and prediction of cognitive performance from actigraphy data.

Methods

General

The study protocol was approved in advance by the U.S. Army Aeromedical Research Laboratory (USAARL) Human Use Committee. The principal investigator (PI) provided detailed briefings to the unit Commanders explaining the purpose, procedures and risks of the study and the actions required of those personnel who might volunteer to participate. Upon receiving the Commanders' approvals, the PI provided each interested Soldier with a comprehensive briefing regarding the study's specifics. Following the presentation, each Soldier met with an ombudsman to clarify and explain any remaining concerns. Those who still wished to participate provided written informed consent.

Study population and description

The study population was composed of 108 military personnel of the age of majority undergoing military training at the Noncommissioned Officer Academy and the Warrant Officer Candidate School (WOCS) at Fort Rucker, AL. Two participants were withdrawn from the study for failing to complete NCO courses, leaving a total of 106. Personnel undergoing official military training must meet the physical and mental criteria for participation in such training. Therefore, there was no need for any research-specific mental or physical evaluation for participation. They were recruited through meetings (prearranged through command channels) prior to the start of their particular training course. The data collected were from a diverse military population (noncommissioned officers and warrant officer candidates; both men and women) and a wide range of training and performance measures.

Although the WOCS and the Advanced and Basic Noncommissioned Officer Courses (ANOC and BNOC, respectively) provide some MOS-specific course content, they are essentially non-technical, leadership courses which provide the training to improve military and professional leadership skills, resource management, and communication skills appropriate to the Soldiers' experience level, rank, and/or future performance expectations. Generally, the training focuses on various aspects of preparing unit and subordinate elements for peace and wartime missions and contingencies; planning, supervising, and executing tasks and missions; leading, supervising, disciplining, training, and developing subordinates; planning, scheduling, supervising, executing, and assessing the unit's mission essential training; planning, initiating, and supervising personnel, administration, and supply actions; planning, supervising, and assessing the safe use, maintenance, storage, security, and accountability of personal and organizational equipment and material; and caring for subordinates and their families (Army Training Requirements and Resources System, 3 Jun 09). In addition, the WOCS provides warrant officer candidates the foundation they need to succeed as warrant officers in the Army and to be adaptable to the ever increasing challenges of the operational environment (U.S. Army Warrant Officer Career College, 28 May 09). All of the NCO and WOCS courses culminate in field training exercises which then applies the flexible, adaptive leadership principles in stressful, sometimes ambiguous, situations to reinforce and build upon previous classroom theory studies and discussion.

There were no gender, age, military occupational skill (MOS), or expertise restrictions and hence, no person of majority age actively participating in the relevant military training was excluded from the study.

As noted, the population was attending military courses specific to certain military career tracks. Table 1 presents the participant population by course and military occupational skill (MOS).

Table 1.
Recruited participants by course and MOS.

<u>Course</u> Military Occupational Skill	Number of Participants	Course Length (days)
<u>Advanced Noncommissioned Officer Course (ANOC) †</u> Air Traffic Controller (15Q40) Common Aviation Maintenance (15[N,R,S,T,U,Z]40) <u>Basic Noncommissioned Officer Course (BNOC) ††</u> Air Traffic Controller (15Q30) Aviation Operations Specialist (15P30)	60	31, 36, 38, or 43*
<u>Warrant Officer Candidate School (WOCS)</u> Warrant Officer Candidates Classes # 17 and 18	48	30

†Sergeants First Class/E-7

†† Staff Sergeants/E-6

* depending upon specific MOS

Measures

Soldier activity data (the predictor variable) was collected through the use of the Actiwatch® activity monitoring system (Mini Mitter Company, Incorporated). According to the Actiwatch® Instructional Manual (2001), the Actiwatch® is a small, lightweight, limb-worn, device which utilizes an accelerometer to monitor the occurrence and degree of motion. The sensor integrates the degree and speed of motion and produces an electrical current that varies in proportional magnitude at a sampling rate of up to 32 Hz. It contains an omnidirectional sensor and is thus sensitive to motion in all directions. Once collected, the data were wirelessly downloaded to a reader which is connected to a personal computer. The accompanying Actiwatch® software allows the manipulation, analysis, and presentation of the output data. The output data (Actiwatch® Instructional Manual, 2001) used in this study's analyses were:

1. Assumed sleep - reflects the amount of time between the point at which it appears that the onset of the sleep period and the time that they finally arose from sleep.
2. Actual wake time - reflects the number of minutes of time within the Assumed Sleep period that were spent awake (i.e., not scored as sleep).
3. Sleep efficiency - indicates the percent of time that the wearer is assumed to be “in bed” or attempting to sleep.
4. Number of sleep bouts - reflects the number of independent bouts of sleep identified during the Assumed Sleep Period.
5. Immobile minutes - reflects the number of minutes within the Assumed Sleep period that were scored as immobile (i.e., without detectible movement).
6. Fragmentation index - reflects the amount of movement or disrupted sleep.

Procedure

Following informed consent, volunteers were assigned a subject number. Subsequently, the PI and research staff issued an Actiwatch® to each participant. Each Actiwatch® was identifiable by serial number and matched to the subject number. The donning of each Actiwatch® on the subject’s wrist was supervised by the research staff to ensure proper placement. The wristband is likened to that of a common two-piece, belt-like, watchband. The volunteer was instructed that the Actiwatch® had to remain on the arm for the duration of the study unless removed for unforeseen health, discomfort, and/or safety reasons. Activity data (the predictor variable) was recorded automatically by the Actiwatch® throughout the duration of the predetermined course-specific period. At course completion, the Actiwatches® were retrieved and the data downloaded. Due to the variability of training syllabi, the exact number of days varied (table 1).

The general design of the study is shown in figure 1. Although the duration of the study differed according to the specific MOS of the volunteer, all groups engaged in a similar process. 1) All participants donned the actigraph and wore it continuously for the duration of the study; 2) each participant completed at least three classroom examinations during the course of the study; and 3) at the completion of the study, each subject returned their actigraph and the data was downloaded to a computer database (figure 1).

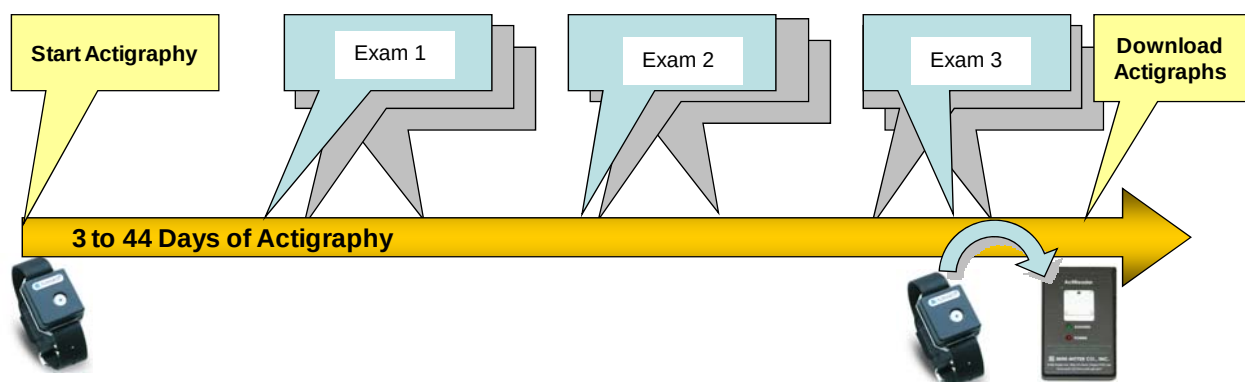


Figure 1. General study design. (For general illustrative purposes only.)

All performance evaluations were conducted by the military course cadre with performance scores passed to the research team upon course completion. Military relevant performance measures (the outcome variables) were those measures/exam scores applicable to the particular course of instruction. Performance measures and the specific times and dates of their collection were amassed for performance-activity comparisons. All identifiable data (name/serial numbers) were converted to subject number reference to ensure subject confidentiality.

Data preparation and analysis

Data retrieval

Retrieval of the Actiwatch[®] from the participants found that three of the 106 devices still issued had been lost during the data collection period thus leaving 103 available for downloading. Each Actiwatch[®] was downloaded using the Actiwatch[®] Reader. Due to unexplained technical difficulties, data were only recorded on 79 Actiwatch[®]s. For each participant, an individual data file (i.e., .AWD file) was downloaded and saved. This file includes a header with the individualized subject number, date and time of actigraph initialization, basic demographics of the participant, and a column of data points representing minute-by-minute activity counts for the duration of the study.

Initial data processing

Three participants admitted to removing the devices during the collection period. (Their data were not included in the final analyses.) Downloading attempts of the actigraphy data further reduced the data pool. For unknown reasons, 21 Actiwatch[®]s failed to download any activity data while 32 collected data for less than 20 days. Participants with data clearly missing were excluded from the data set. The remaining participants with initially valid appearing actigraphy data were then entered into an Excel spreadsheet. All examination scores for these participants were entered into this sheet as well.

Overall analysis plans

Three approaches were undertaken to analyze the data: Approach 1) Use only Actiwatch[®] sleep scores from the two days preceding the exam to predict performance on the exam (figure 2); Approach 2) Use only Actiwatch[®] sleep scores from one day preceding the exam to predict exam performance (figure 3); Approach 3) Use the Sleep History and Readiness Predictor (SHARP) to predict exam performances based on a previously validated complex mathematical prediction model that accounts for all days of measured sleep preceding the exam and current time of day to predict exam performance (figure 4).

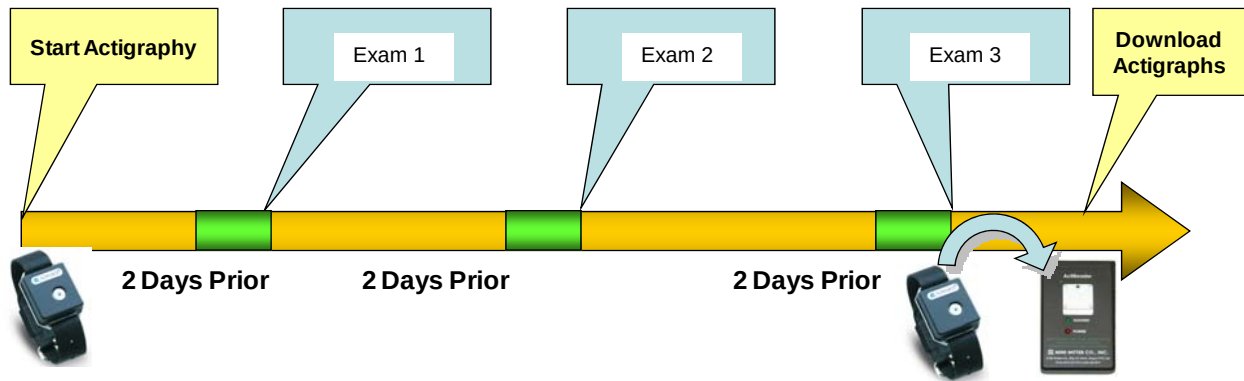


Figure 2. Using average measured sleep during the two days preceding each examination to predict exam performances. (For general illustrative purposes only.)

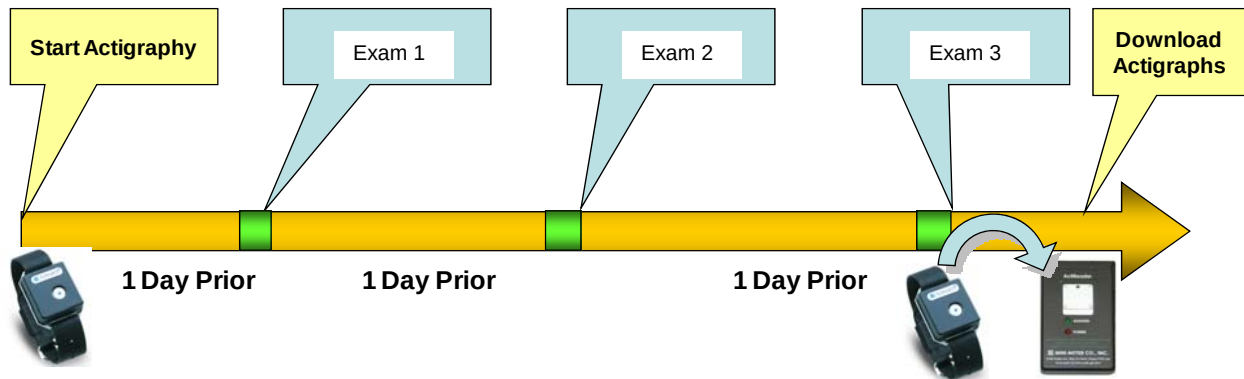


Figure 3. Using the average measured sleep during the one day preceding each examination to predict exam performances. (For general illustrative purposes only.)

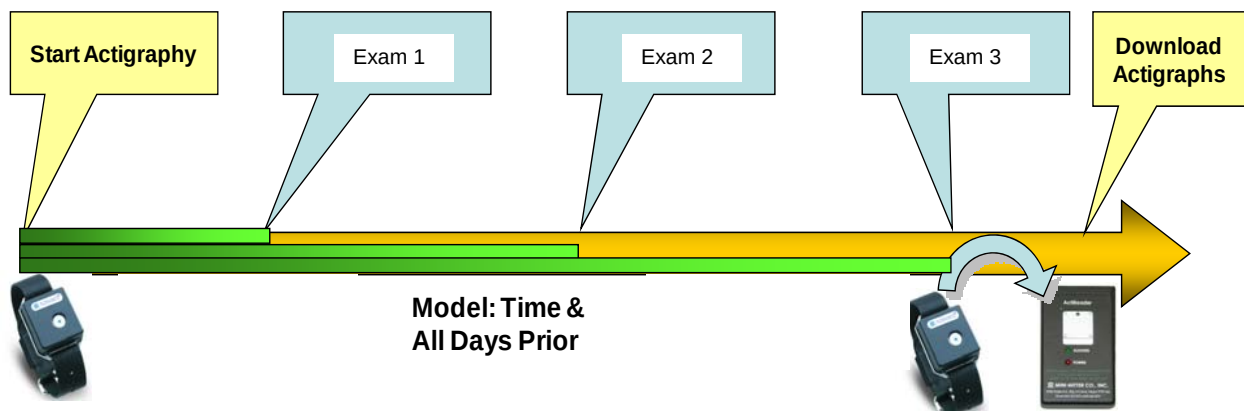


Figure 4. Using the SHARP data analysis to examine the ability of the sleep-performance prediction model for all days preceding each examination to predict exam performances. (For general illustrative purposes only.)

Actiware analysis

Overview

Initially, it was of interest to extract raw sleep scores (i.e., independent of model predictions) and evaluate whether they were predictive of militarily relevant cognitive performance. For the present study, three parameters were analyzed: 1) the relationship between total average sleep and performance; 2) the relationship between the average amount of sleep obtained in the 48 hours (hr) preceding an exam and exam performance; and 3) the relationship between the average amount of sleep obtained in the 24 hours preceding an exam and exam performance. Actiware Analysis followed three stages: 1) for every subject, Actiwatch data were extracted for each day and analyzed using the Actiware 3.41 Sleep Analysis program; 2) data for each day were validated by visual inspection and a computerized algorithm to ensure that they reflected true sleep/wake measurements; and 3) data were averaged for time periods of interest (i.e., all days of the study; two days before each exam; one day before each exam).

Data extraction and sleep analysis

For each subject, the actigraph data file (i.e., .AWD file) was individually opened using Actiware 3.41. Using the Sleep Analysis feature of the program, the analysis period was set to begin at 21:00 and end at 11:00 the following day (figure 5). This was set for each day (i.e., 24-hour period) of data. By implementing the “Auto” function of the Actiware 3.41 program, estimated sleep onset and wake times as well as a series of sleep statistics were calculated for each day within the record and output as separate data columns for each day. By default, the program calculates statistics on 23 sleep-related variables, including Actual Sleep Time, Sleep Efficiency, Sleep Latency, and Fragmentation Index, among others. Data columns for each 24-hour period were extracted and entered into the study spreadsheet for further analysis. An example of the data that were extracted is shown in figure 6.

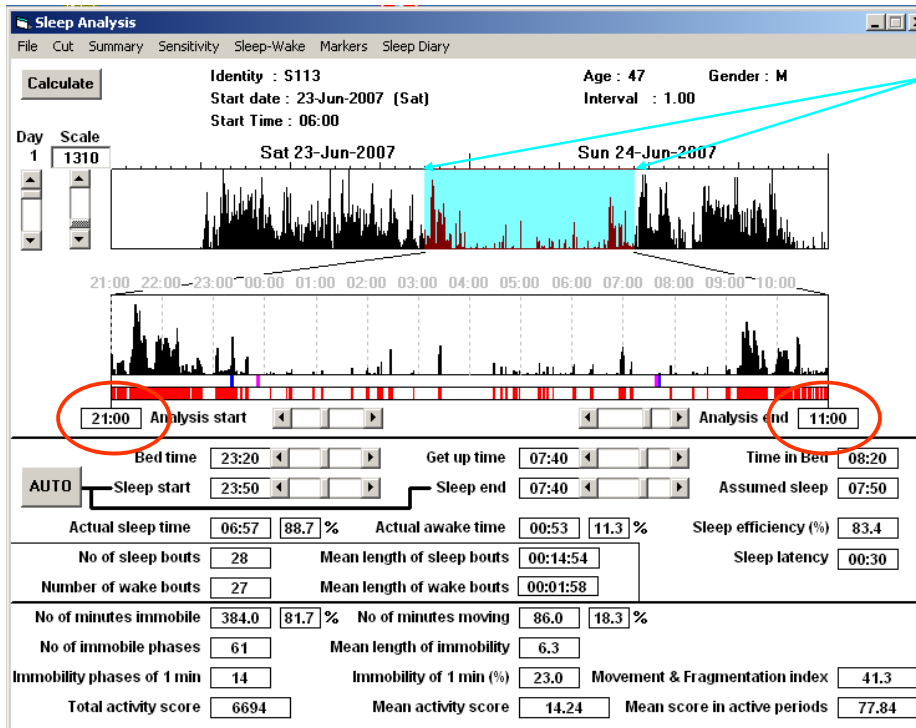


Figure 5. Actiware 3.41 Sleep Analysis window.

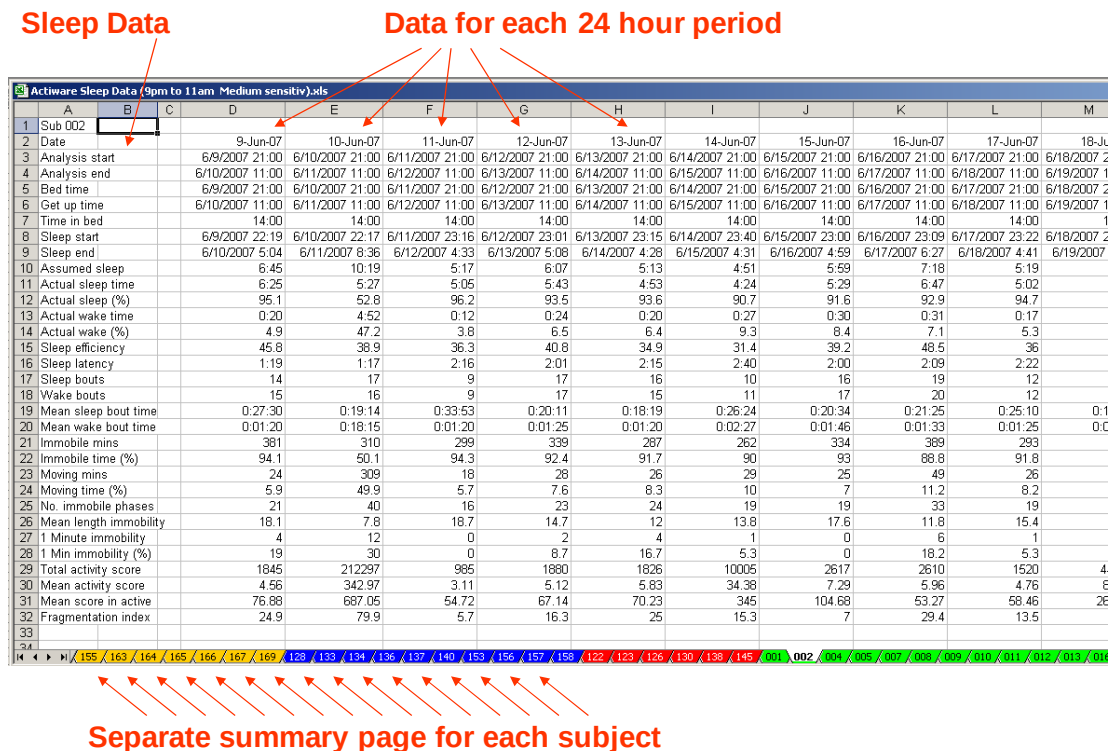


Figure 6. Example of the daily sleep scores extracted for one subject.

Data quality check

Visual quality inspection of the data

As a second stage of data cleaning, a sample of data files was opened in the Actiware® actigraphy analysis program version 3.41 and inspected for quality by a trained and experienced sleep scientist. This involved visually inspecting the “actigram” of each day of collected data individually for every subject in the study, for sample of 587 (30 percent) recorded days out of the entire set of 1945 recorded days. During visual inspection, data for each day were coded as “Valid,” “Probably Invalid,” or “Definitely Invalid.” Figure 7 shows examples of prototypical actigrams for each of these three categories. The goal was to remove data that were judged as unlikely to be valid measurements of continuous human activity. For example, a subject might have removed their Actiwatch temporarily and neglected to put it back on within a short period of time. This would result in a large period of blank readings that would be incorrectly scored as sleep by the automated scoring of the Actiware program. Therefore, such invalid days would need to be excluded from analysis, or they could severely skew the results. In this case, the visual inspection was used to identify records that were clearly atypical (e.g., zero activity for many continuous hours), and remove those from the dataset.

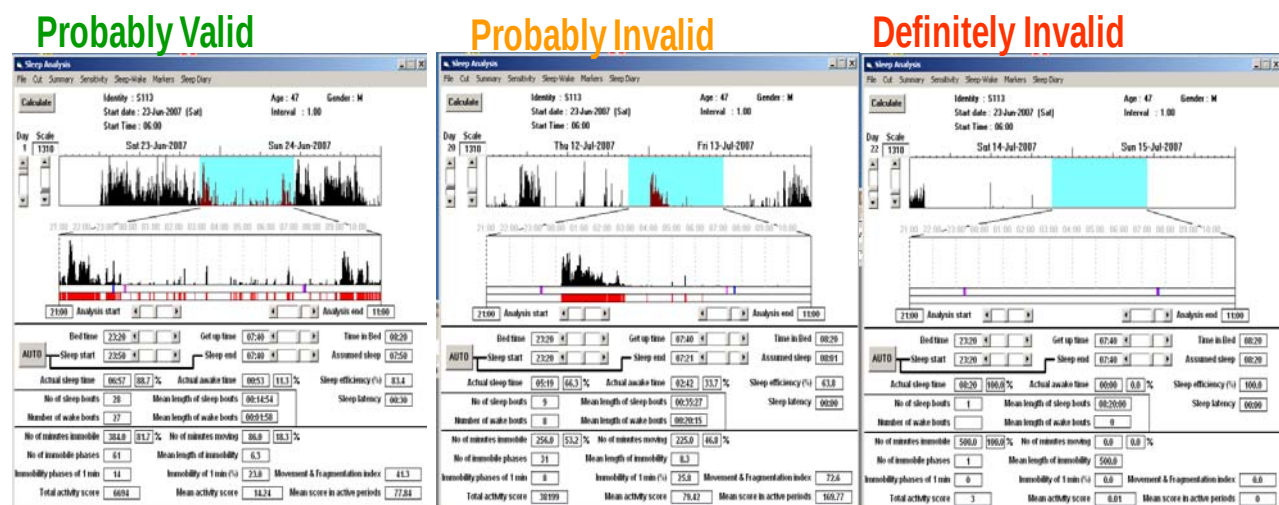


Figure 7. Examples of valid, probably invalid, and definitely invalid actigrams by visual inspection.

Automated quality inspection of the data

Because the quality inspection process was subjective and required visual examination of each day of each record by an experienced sleep researcher, an attempt to quantify and automate this process was also undertaken. A computer algorithm was written as a macro in Microsoft Excel to determine whether the findings from the visual quality inspection of the data could be automated for consistency within and across future studies. To develop this algorithm, datasets identified as valid and invalid through visual inspection by the experienced sleep researcher were examined for potential differentiating variables. Specifically, each of the output variables

provided by the Actiware program was compared between the valid and invalid profiles. Based on this examination, eight variables were identified as distinguishing between those that appeared valid and those that appeared invalid. These criteria were empirically derived and based solely upon their ability to differentiate between apparently valid and invalid datasets from the current sample. Based on this examination, the eight variables showing the greatest utility in discriminating between valid and invalid days of actigraphy were:

1. Actual Sleep Time > 10 hs
2. Sleep Efficiency > 60 percent
3. Sleep Latency < 1 minute (min)
4. Number of Sleep Bouts < 10
5. Mean Sleep Bout Time > 2.5 hr
6. Immobile Minutes > 600
7. Immobile Phases < 12
8. Mean Length of Immobility > 100 min

Subjects with valid appearing data were unlikely to score positive on few, if any, of these criteria. However, a positive score on any one of these criteria alone was not enough to identify a sleep record as invalid. It was assumed, however, that because each of these criteria had a low probability of occurrence, a combination of several items meeting these criteria in the same record would raise further concern of potential validity issues. Thus, each item was scored as “0” if the criterion was not met and “1” if the criterion had been exceeded. The scores for these eight items were summed to provide a total validity score. Higher scores on this index (i.e., greater number of atypical scores) indicated a greater likelihood that the record was valid.

To determine the ideal cut-off for identifying a record as valid or invalid, classifications using the obtained index scores were compared to the validity determinations made by an experienced sleep researcher. The percent agreement and Kappa coefficients for various cut-off scores were evaluated statistically as depicted in table 2.

Table 2.
Kappa coefficients.

Cut-Off Value	Kappa Coefficient	Percent Agreement
2	94.2	97.3
3	96.0	98.1
4	93.3	96.9

As shown in table 2, a cut-off value of 2 positive items on the index resulted in agreement with the expert rater 97.3% of the time, while a cut-off value of 3 resulted in a 98.1% agreement. Based on these statistics, it was apparent that a cut-off value of ≥ 3 provided the best agreement with the expert visual inspection. Thus validity index scores of 3 or greater were used to eliminate data columns that might represent off-wrist or other data collection problems. This algorithm was run on the entire set of 1945 recorded days. Any column of data meeting this ≥ 3 criterion was excluded from the analysis. All valid data for each subject were imported from the excel spreadsheet into SPSS for further statistical analysis.

Final sample after data cleaning

Out of the 1945 recorded days, 1463 (75.2%) were considered valid while 482 (24.8%) days were coded as invalid. On average, each participant had 30.4 days ($SD = 3.2$) of measured data and 22.9 days ($SD = 8.2$) of valid data. After accounting for data loss due to operational failures of the Actiwatch[®], failures of the participants to follow study instructions, and data cleaning, a final sample of 64 participants with valid data was retained. The field conditions of the study did not allow for 100 percent supervision of the participants by the research staff. During the data collection periods, the PI visited the WOCS and the Noncommissioned Officer Academy at least once a week to ensure there were no reports of problems or issues with participants relating to the Actiwatch[®] data collection. No problems or issues were ever reported. Despite the recurring spot checks by the PI, it is suspected that some participants, in addition to the admitted three, may have removed the devices and stored them away for extended periods until the time of retrieval. That would account for cases in which the downloaded data indicated several days of inactivity. That said, other than the three subjects mentioned above, no other participants reported or were ever witnessed removing their devices during the data collection period.

Output

As mentioned above, the Actiware program provides a large number of summary sleep variables, 23 of which were included in the present study (e.g., Actual Sleep Time, Sleep Efficiency, Sleep Latency, and Fragmentation Index, etc.). These summary variables were calculated for three different time periods: 1) for the two days preceding each exam; 2) for the one day preceding each exam; and 3) for all days of measured actigraphy prior to each exam.

SHARP analysis

Procedure. Data files for each academy class were analyzed using the WRAIR Sleep History and Readiness Predictor (SHARP) program. This program provides a simple graphical user interface to analyze actigraphy data and make predictions about cognitive exam performances during pre-specified time periods or “Mission Windows.” As described previously, the SHARP uses the SAFTE sleep-performance prediction model and the computing machinery of the FAST to provide the mathematical calculations. For the present study, this was accomplished by 1) activating the SHARP program, 2) selecting all relevant actigraphy data files (i.e., .AWD files) for a specific academy class, 3) entering the date and start and end times of the course examination of interest into the appropriate boxes on the SHARP graphical user interface, 4) entering a time resolution or “smoothing” value for the data, and 5) pressing the “Analyze Files” button (figure 2).

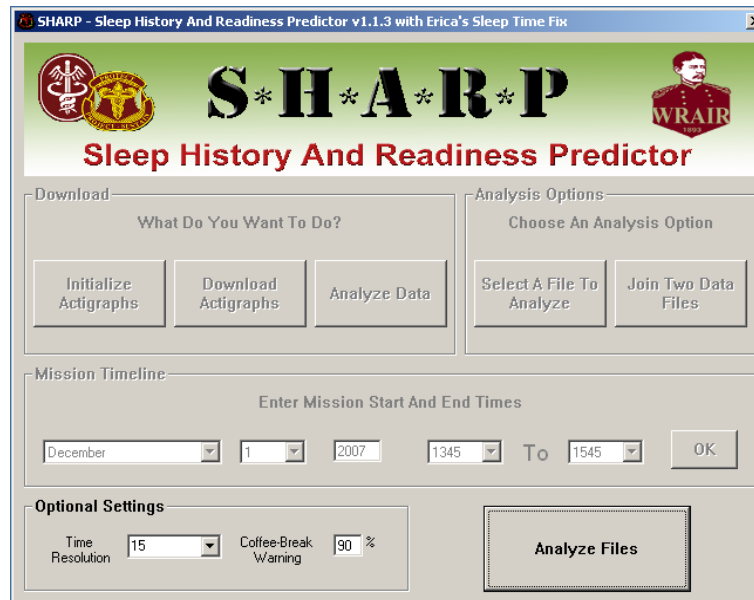


Figure 8. SHARP graphical user interface.

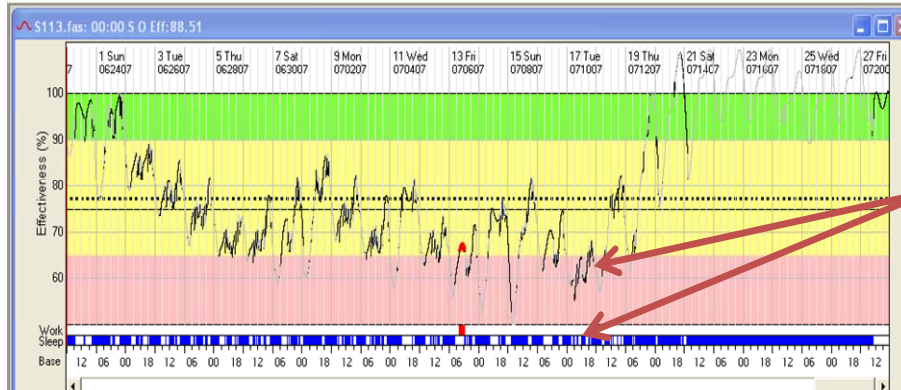
SHARP output. All subjects within a specific class of the ANOC, BNOC, and WOCS courses were processed together as a group because they all started the study together and took the examinations at the same times. At the completion of each analysis run, the SHARP produced a summary output table in Microsoft Excel (table 3). This table included a listing of subject identifiers (i.e., subject number), a summary score indicating the mean FAST Effectiveness Score (i.e., Mission Effectiveness [ME]) during the time of the examination being evaluated (i.e., Mission Window), a score indicating the percent of the Mission Window that falls below a criterion ME score of 90 (i.e., Percent Below Criterion), an index of the mean FAST Effectiveness scores during all other waking times not included in the current Mission Window (i.e., Typical Awake Effectiveness score), an estimate of the mean hours of sleep obtained during the entire time that the actigraph has been worn (i.e., Average Daily Sleep), the length of the examining period (i.e., Mission Length), the number of days of actigraphy data that were measured to obtain the current output (i.e., Days Measured), and the number of hours between the time the exam was taken and the time that the actigraphs were downloaded (i.e., Time Since Download—negative values indicate that the exam day occurred before the download, positive values mean that the program is predicting scores after download and may have some error due to activity/sleep unaccounted for after download). A similar table was generated for each of the three examination periods for each academy class that participated in the present study. These tables were then pasted into the previously described spreadsheet that included all of the examination scores.

Table 3.
SHARP output table.

Name	Mission Effective	Percent Below Criterion	Typical Awake Effective	Average Daily Sleep	Mission Length	Days Meas.	Time Since Dwnld
S001	99	0	93	13h 33m	1h 0m	31	-49h -48m
S002	97	0	90	7h 5m	1h 0m	31	-50h -16m
S004	93	0	87	6h 18m	1h 0m	31	-50h -19m
S005	88	100	90	7h 58m	1h 0m	31	-48h -56m
S007	101	0	91	12h 16m	1h 0m	31	-49h -26m
S008	87	100	89	6h 35m	1h 0m	31	-49h -7m
S009	91	0	89	6h 47m	1h 0m	31	-344h -53m
S010	71	100	83	6h 29m	1h 0m	31	-44h -11m
S011	85	100	89	10h 45m	1h 0m	31	-49h -13m
S012	102	0	87	11h 22m	1h 0m	31	-49h -33m
S013	96	0	91	7h 13m	1h 0m	31	-50h -8m
S016	87	100	90	6h 36m	1h 0m	31	-48h -58m
S017	99	0	92	13h 15m	1h 0m	31	-50h -22m
S018	90	38	89	6h 35m	1h 0m	31	-172h -19m
S019	88	100	91	6h 48m	1h 0m	31	-50h -4m
S020	95	0	88	6h 43m	1h 0m	31	-49h -54m
S022	94	0	89	7h 11m	1h 0m	31	-50h -11m
S024	97	0	91	7h 31m	1h 0m	31	-50h -26m
S025	90	67	87	6h 9m	1h 0m	31	-49h -36m
S026	89	100	89	7h 3m	1h 0m	31	-49h -21m

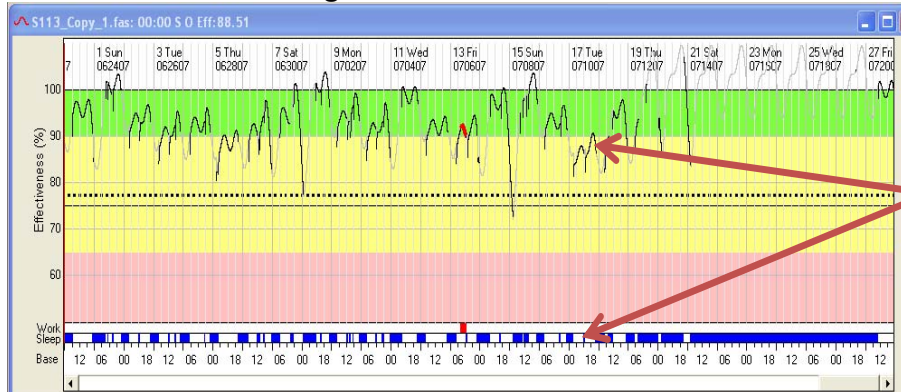
Smoothing iterations. The analysis process was repeated three times, each with a different “time resolution” (i.e., smoothness). The time resolution of the data refers to a function within the FAST program that allows the user to specify the stringency of the criteria for classifying whether a specific time period is sleep or wake. The FAST allows the user to enter a value ranging from “0” to “30” min, which will be averaged into a single “block” of time and scored in a binary fashion as either sleep or wake. Larger values require a longer duration of sustained absence of movement in order for that particular time block to be classified as a period of sleep. Thus, a time resolution of “0” would mean that any minute scored as “sleep” by the actigraph would be analyzed as a sleep period by the FAST program. For example, an hour would be made up of 60 one-minute periods, each of which would be scored as sleep or wake. This could be problematic in some situations, as a moment of inactivity lasting 60 s would be interpreted by the program as actual sleep when in fact the subject may have been sitting still. A longer time resolution of “15” would provide more protection against such spurious inflation of sleep scores by evaluating every 15-min block and determining whether the majority of it was scored as wake or sleep and then coding the entire block according to the majority values (e.g., 8 min of sleep would code the entire 15-min period as sleep; 7 min of sleep would code the entire block as awake). Similarly, a smoothing value of “30” would evaluate each 30-min block of time and re-code it based on the whether the majority was wake or sleep. An example of the effect of no smoothing versus 30-min smoothing on actual data is shown in figure 9. To evaluate the validity of this “smoothing” procedure, the data for each subject were processed at three time resolutions: 0, 15, and 30 minutes. The resulting data tables were each incorporated into the larger spreadsheet for statistical analysis.

No Smoothing



With no smoothing, many time bins are counted as “sleep” even when subject was actually awake

30 Minute Smoothing



With 30 minute smoothing, many of the same time bins are correctly counted as “wake”

Figure 9. Example of the two smoothing levels on the data.

Results

Statistical analyses

All statistical analyses were conducted using SPSS® 12.0 with statistical significance set at an alpha level of .05 for all statistical exams.

Actiware analyses

Average total sleep per night

From the valid Actiware scores, the average sleep per night was calculated for the sample as a whole ($n = 64$). The mean sleep obtained per night across all classes evaluated was 5.80 hr ($SD = 0.55$). Figure 10 shows the distribution of average nightly sleep for the entire sample, which ranged from a minimum of 4.5 hr to a maximum of 7.5 hr per night. The National Sleep Foundation (14 Sep 07) recommends that adults obtain 7 to 9 hr of sleep per night. In the present sample, only three Soldiers (3.1 percent) met this criterion of adequate sleep.

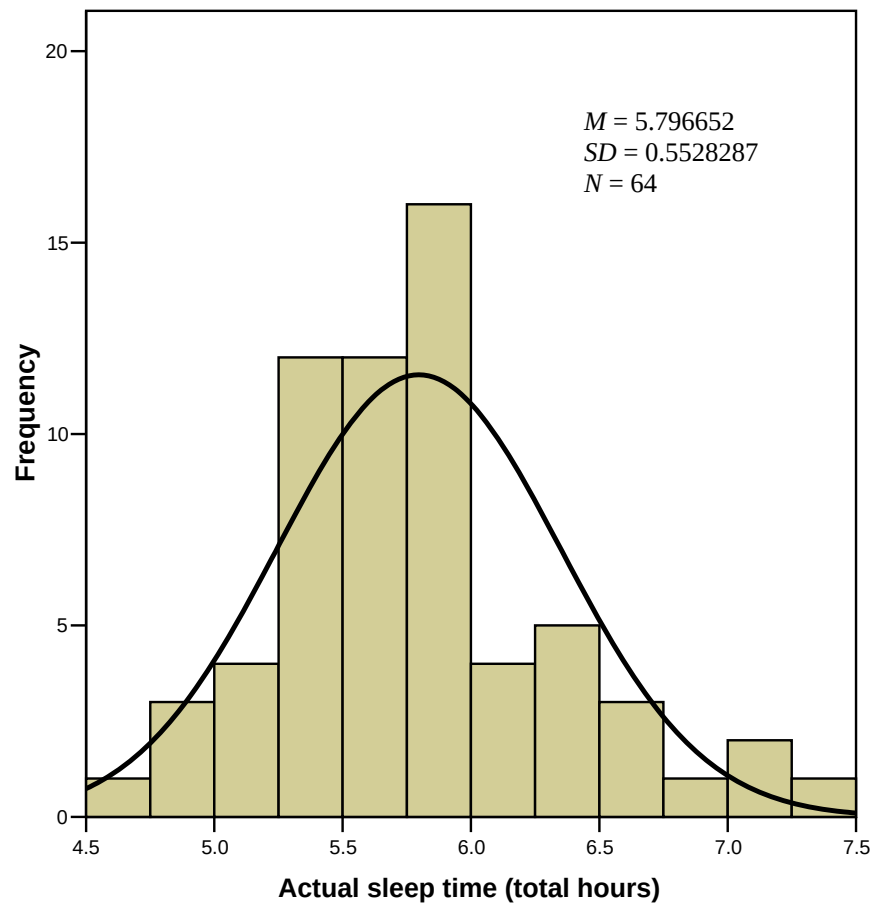


Figure 10. Histogram of the mean hours of sleep obtained by all participants for the entire duration of the study.

Comparisons across class groups

Total sleep time

It was also of interest to examine whether there were significant differences among the various courses in the average amount of sleep obtained by the class members. A one-way analysis of variance (ANOVA) showed that there was, in fact, a significant main effect of course on the amount of sleep obtained, $F(5,58) = 3.09$, $p = .015$. The average sleep obtained by each class is depicted in figure 11. Post-hoc pair-wise comparisons (LSD exams) showed that the Maintenance Course and the 15Q40 Air Traffic Control Course generally obtained the most sleep whereas the 15P30 Aviation Operations Specialist Course and both WOCS classes obtained the least. Significant ($p < .05$) pair-wise differences are shown by the red brackets in figure 11. Complete data for total sleep time in all groups can be found in table 4.

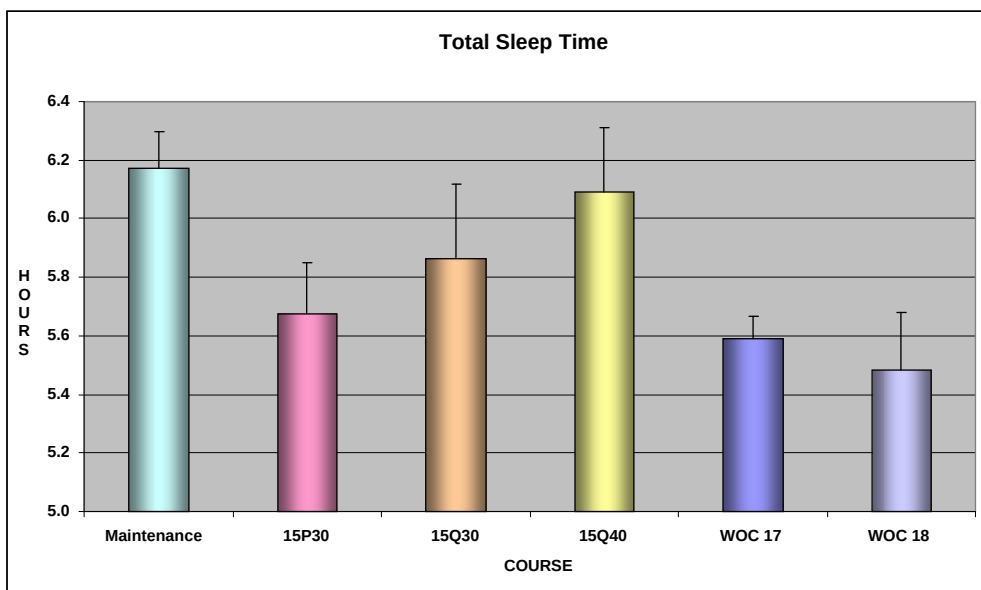


Figure 11. Mean total sleep durations for each course group (significant differences exceeding $p < .05$ for LSD exam are indicated by brackets).

Table 4.
Group statistics for total sleep time in hours.

Group	Mean	SD	SE	95%CI Lower	95% CI Upper	Min	Max
Maintenance (n=13)	6.170	0.451	0.125	5.897	6.443	5.595	7.037
15P30 (n=8)	5.677	0.491	0.173	5.266	6.087	4.921	6.458
15Q30 (n=10)	5.863	0.805	0.254	5.287	6.439	4.936	7.468
15Q40 (n=6)	6.093	0.529	0.216	5.537	6.649	5.374	6.738
WOC 17 (n=20)	5.589	0.338	0.075	5.430	5.747	4.598	5.981
WOC 18 (n=7)	5.482	0.523	0.197	4.998	5.966	4.860	6.505

Sleep efficiency

Similar to the findings for total sleep time, mean scores for sleep efficiency, a measure of the quality of sleep obtained, showed a significant difference across the classes, $F(5, 58) = 3.68, p = .006$. Sleep efficiency reflects the percent of time that a Soldier is assumed to be attempting to sleep that is actually spent asleep. In other words, it is the ratio of time spent asleep to the amount of time in bed. A Soldier who spends 8 hours in bed, but sleeps for only 6 hours would have a sleep efficiency of 75%. The pattern of mean sleep efficiency scores was very similar to that found for total sleep, as evidenced by the post-hoc comparisons between means (figure 12). On average, the mean sleep efficiency across the entire sample was 41.32 percent ($SD = 3.91$). Complete data for sleep efficiency in all groups can be found in table 5.

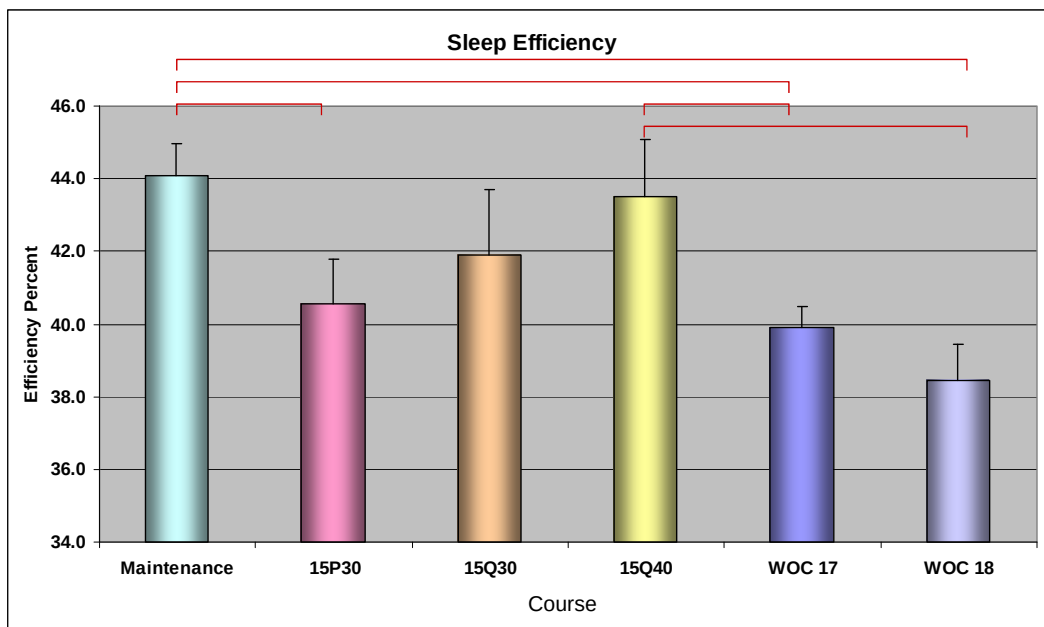


Figure 12. Mean sleep efficiency (in percent) for each course group (significant differences exceeding $p < .05$ for LSD exam are indicated by brackets).

Table 5.
Group statistics for sleep efficiency percent.

Course	<i>M</i>	<i>SD</i>	<i>SE</i>	95%CI Lower	95% CI Upper	Min	Max
Maintenance (<i>n</i> =13)	44.073	3.227	.895	42.123	46.023	39.970	50.275
15P30 (<i>n</i> =8)	40.548	3.513	1.242	37.611	43.486	35.142	46.144
15Q30 (<i>n</i> =10)	41.879	5.753	1.819	37.763	45.995	35.250	53.344
15Q40 (<i>n</i> =6)	43.522	3.783	1.544	39.551	47.493	38.390	48.133
WOC 17 (<i>n</i> =20)	39.922	2.421	.541	38.788	41.055	32.834	42.718
WOC 18 (<i>n</i> =7)	38.449	2.610	.986	36.035	40.863	34.493	42.462

Sleep latency

The index of sleep latency (i.e., the amount of time to fall asleep once in bed) was also found to differ significantly among the course groups, $F(5,58) = 6.77, p < .001$. Data in figure 13 indicate that all groups showed some significant difficulty falling asleep, with the sample as a whole taking 2.05 hr ($SD = 0.85$) to fall asleep on average. Post-hoc comparisons among course groups suggested that the 15Q30 and 15P30 courses had the longest sleep onset latencies, whereas the participant in the two Warrant Officer courses tended to fall asleep the fastest. Complete data for sleep latency in all groups can be found in table 6.

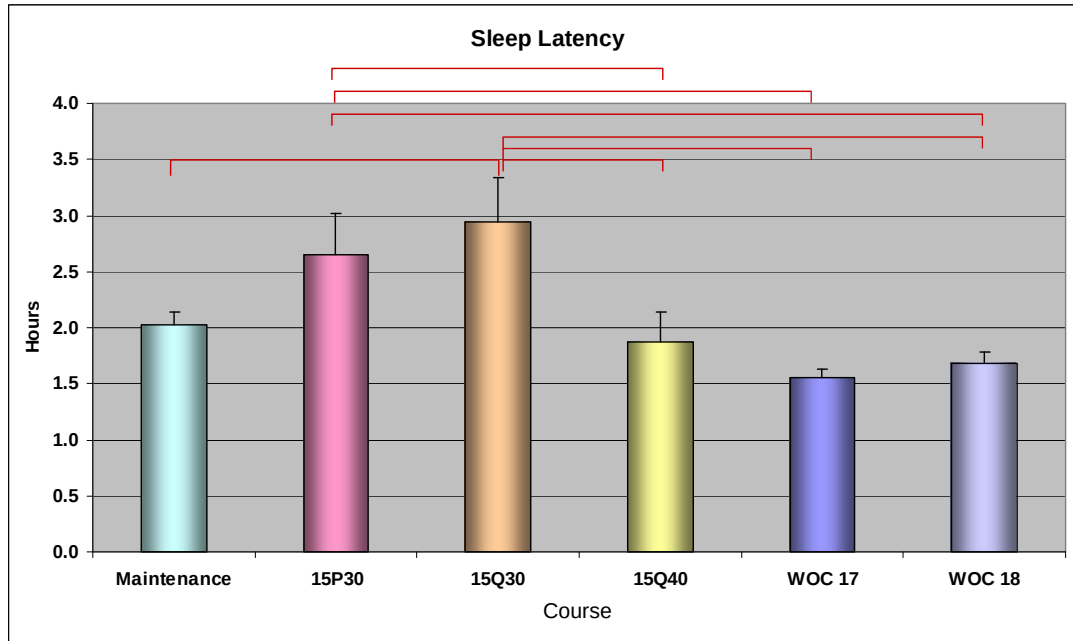


Figure 13. Mean Sleep Latency for each course group (significant differences exceeding $p < .05$ for LSD exam are indicated by brackets).

Table 6.
Group statistics for sleep latency in hours.

Course	<i>M</i>	<i>SD</i>	<i>SE</i>	95%CI Lower	95% CI Upper	Min	Max
Maintenance (<i>n</i> =13)	2.021	.439	.121	1.755	2.286	1.115	2.471
15P30 (<i>n</i> =8)	2.650	1.041	.368	1.779	3.520	1.038	4.400
15Q30 (<i>n</i> =10)	2.937	1.258	.397	2.037	3.837	1.301	4.562
15Q40 (<i>n</i> =6)	1.873	.638	.260	1.202	2.543	1.126	2.609
WOC 17 (<i>n</i> =20)	1.555	.336	.075	1.397	1.713	1.009	2.418
WOC 18 (<i>n</i> =7)	1.687	.255	.096	1.451	1.923	1.143	1.939

Sleep fragmentation

There was also a significant main effect of course on the sleep fragmentation index, which measures the quality of obtained sleep as a function of the frequency of waking throughout the sleep period, $F(5,58) = 2.83, p = .024$. On average, the sample demonstrated a mean Fragmentation Index of 38.03 ($SD = 9.97$). As evident in figure 14, post-hoc comparisons showed that this effect was driven primarily by the WOC 17 group, which obtained a significantly lower sleep fragmentation index than either the 15Q30 or the 15Q40 courses. complete data for sleep fragmentation in all groups can be found in table 7.

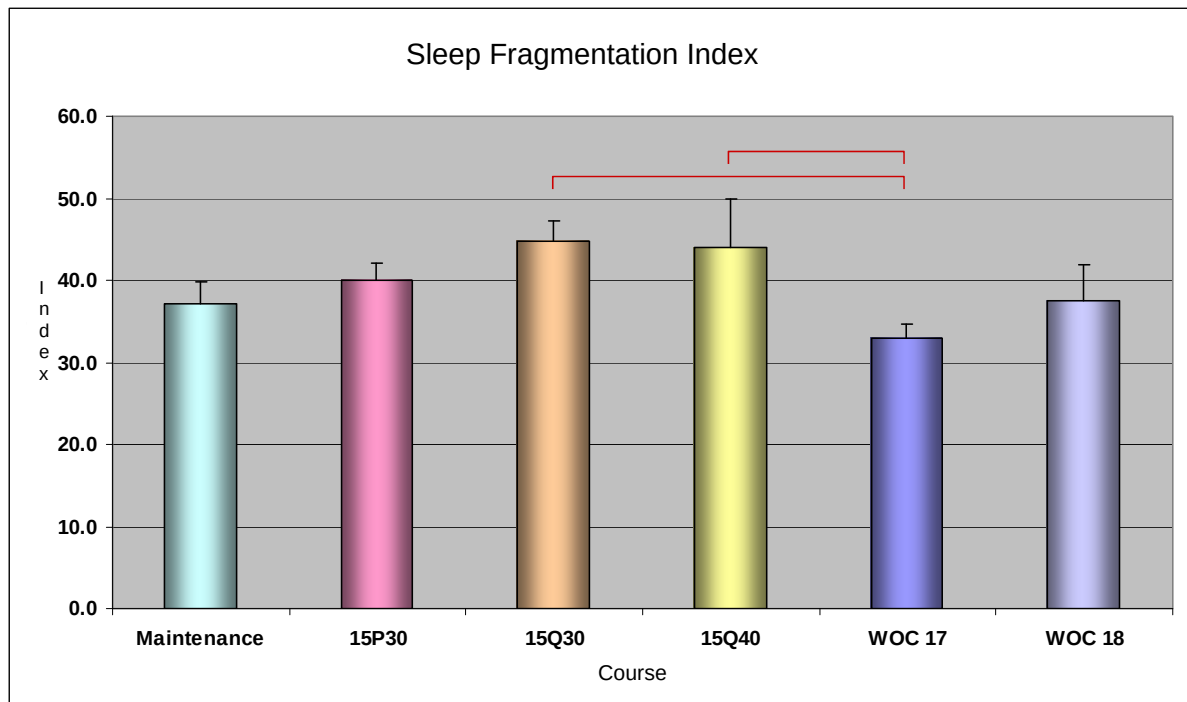


Figure 14. Mean sleep fragmentation index for each course group (significant differences exceeding $p < .05$ for LSD exam are indicated by brackets).

Table 7.
Group statistics for sleep fragmentation index.

Course	<i>M</i>	<i>SD</i>	<i>SE</i>	95%CI Lower	95% CI Upper	Min	Max
Maintenance (<i>n</i> =13)	37.085	9.856	2.733	31.129	43.041	25.842	52.860
15P30 (<i>n</i> =8)	39.907	6.416	2.268	34.543	45.272	33.100	50.244
15Q30 (<i>n</i> =10)	44.711	8.104	2.562	38.914	50.509	28.284	59.861
15Q40 (<i>n</i> =6)	43.994	14.522	5.928	28.754	59.235	26.744	59.816
WOC 17 (<i>n</i> =20)	32.905	7.867	1.759	29.223	36.588	19.658	53.293
WOC 18 (<i>n</i> =7)	37.576	11.209	4.236	27.209	47.943	28.990	54.206

Prediction of exam performance from actigraphy

Overview

A goal of this effort was to evaluate the effectiveness of actigraphy as a predictor of cognitive performance under non-laboratory conditions. For the present study, the dependent variable was performance on the first examination in each course. This examination differed across courses and did not necessarily contain the same content. However, examination scores were converted to percent correct for analysis, so that they were all on a similar scale. Due to small samples within each class, the subjects from all different classes and courses were combined into a single sample, despite differences in course content. Recall that all of the courses were essentially non-technical, leadership courses which provided the training to improve military and professional leadership skills, resource management, and communication skills appropriate to the Soldiers' experience level, rank, and/or future performance expectations. By combining scores across courses, it was possible to obtain a conservative and broadly generalizable assessment of the effects of sleep on cognitive performance (i.e., the procedure tested the effects of sleep/wake schedules on "exam performance" broadly defined, rather than on a specific course content). Although it can be argued that the scores used were generated using several different examinations, the content of which varied, the examinations were appropriate to address and challenge a specific level of Soldier experience and performance expectations. Therefore, for the purposes of this analysis, performance (examination scores) was used as a relative factor regardless of examination content. Several indices of sleep quality were used to predict performance on these examinations across all participants in the study. In order to ensure that valid data were obtained for prediction, we only selected cases that had at least five days of valid data prior to the examination.

Total sleep prior to exam 1

It was of interest to examine the amount of sleep obtained during the period encompassing the two-day period prior to Exam 1 as well as one-day prior to exam 1. By comparing these indices to overall sleep during the course, it was possible to evaluate whether subjects obtained less sleep in the day or two prior to taking an exam (e.g., staying up late "cramming" for the exam). For this analysis, 38 participants had valid data for the five days prior to exam 1. From the Actiware analyses it was determined that the mean hours of sleep obtained over the two-day period prior to the first exam was 4.83 ($SD = 0.74$). A paired-samples t -exam revealed that this amount of sleep over that two-day period was significantly lower than the average sleep obtained by this subsample ($M = 5.79$, $SD = 0.47$) over the duration of the entire study, $t(37) = 9.69$, $p < .001$, suggesting that subjects were altering their normal sleep in the two-day period prior to the exam. Similarly, when the analysis was restricted to one day prior to the first exam, participants obtained a mean of 4.62 ($SD = 1.15$) hr of sleep, which was significantly less than they had obtained two nights before the exam, $t(37) = 2.07$, $p = .046$. This suggests that normal sleep patterns were significantly reduced the night before an exam (figure 15).

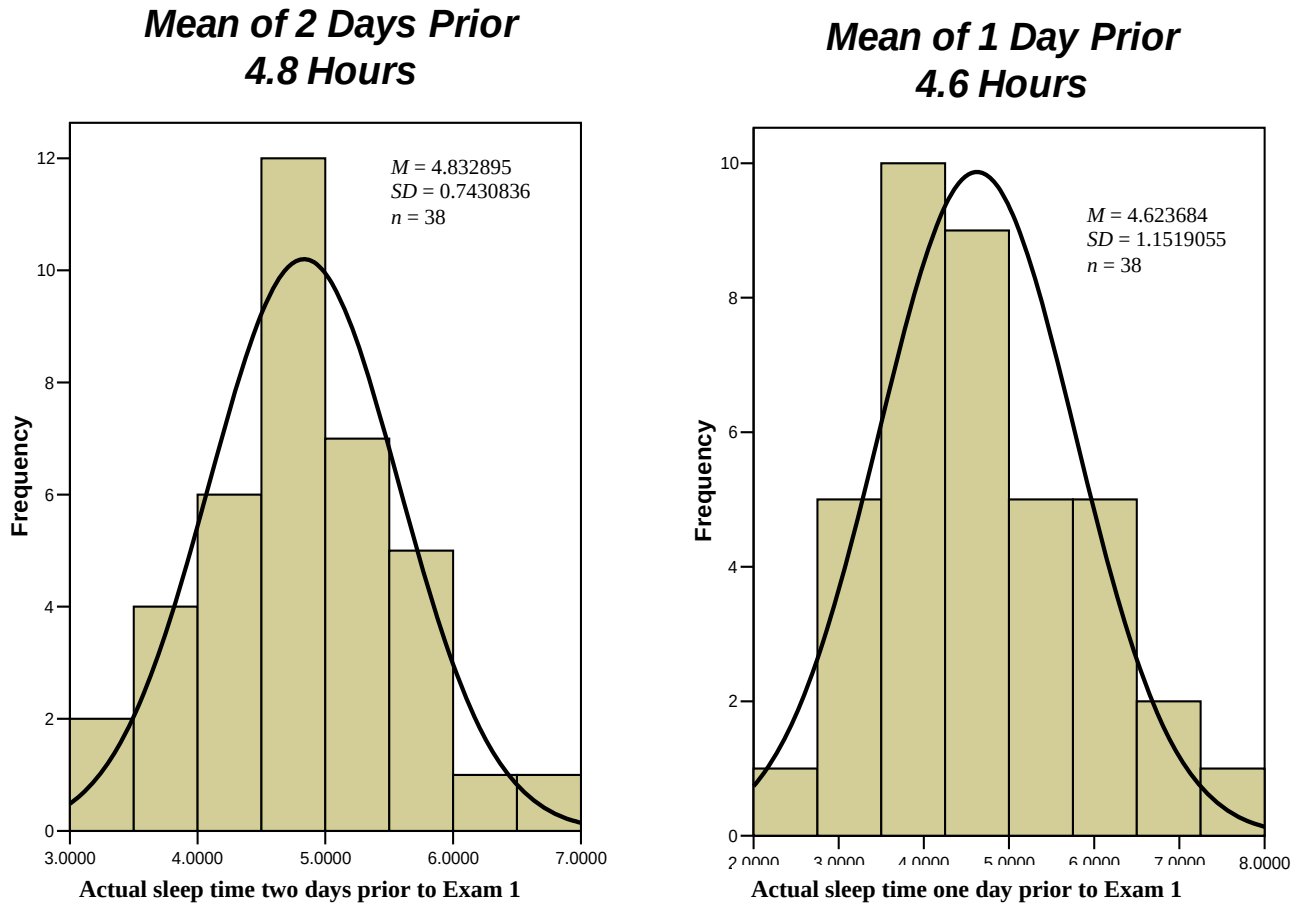


Figure 15. Histograms showing the amounts of sleep obtained two days and one day prior to Exam 1. Note: Selected only subjects with $\geq$ five days of valid data ($n = 37$ for Exam 1).

Relationship between total sleep and performance

Given that some Soldiers altered their normal sleep prior to the exam period, it was of interest to predict their performance on the exam based on the amount of sleep that had been obtained in the preceding nights. Actigraphically measured sleep was entered into a correlation analysis with performance on Exam 1. As evident in figure 16, the average sleep obtained during the two days preceding the exam was significantly and positively correlated with exam performance ($r = .60$, $p < .001$), despite the fact that the exams were from different courses and covered different content. Similarly, when only sleep from the one day preceding the exam was evaluated, there was still a highly significant positive correlation between the amount of sleep obtained and exam scores ($r = .54$, $p < .001$; figure 16). These findings suggest that those Soldiers that obtained the most sleep during the two nights preceding the exam performed significantly better than Soldiers that obtained less sleep.

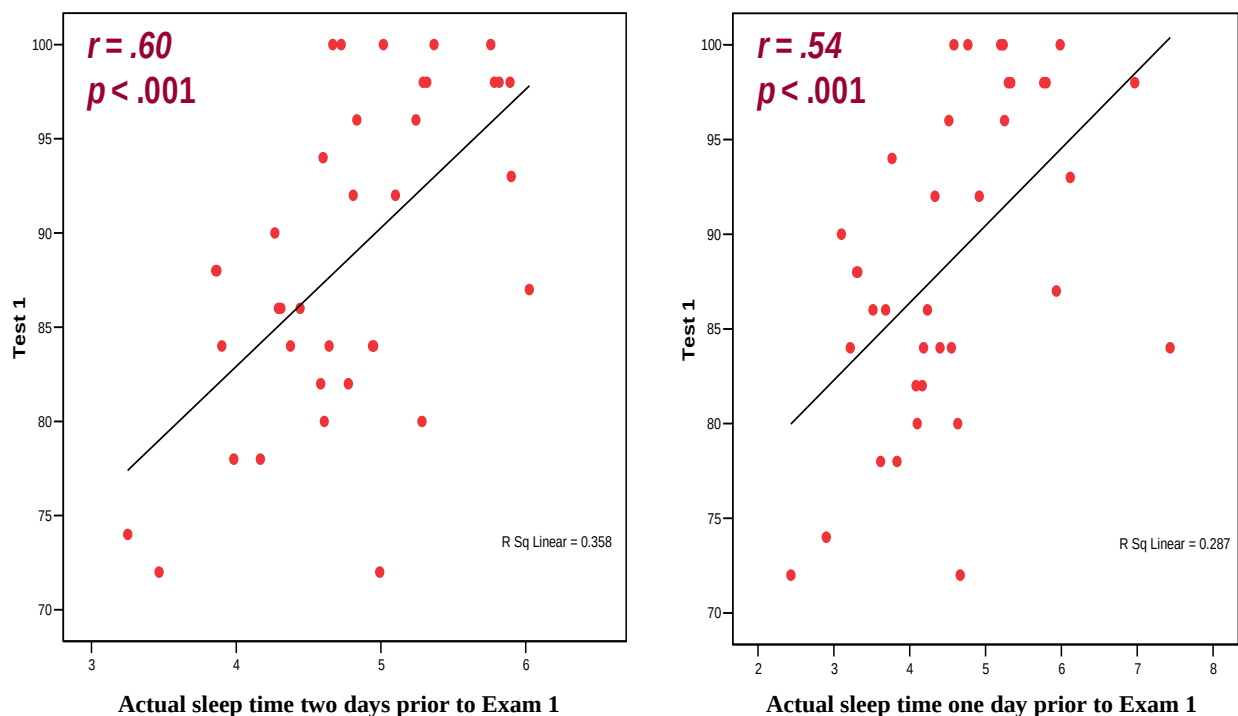


Figure 16. Scatterplots showing the significant positive correlation between exam scores and the amount of sleep obtained two days and one day prior to Exam 1. Note: Selected only subjects with $\geq$ five days of valid data ($n = 37$ for Exam 1).

Relationship between sleep latency and performance

Sleep latency reflects the amount of time required to fall asleep once it is determined that an individual Soldier was probably in bed and attempting sleep. For every Soldier, this index was calculated according to the standard algorithm provided by the Actiware program for each night of sleep. Sleep latency scores were entered into a correlation analysis with performance on Exam 1. As evident in figure 17, the average sleep latency during the two days preceding Exam 1 was significantly and negatively correlated with exam performance ($r = -.46$, $p = .002$). Nearly identical results were obtained when only sleep from the one day preceding the exam was evaluated, with a highly significant negative correlation between sleep latency and exam scores ($r = -.46$, $p = .002$; figure 17). These results suggest that those Soldiers that had the greatest difficulty falling asleep during the two nights preceding the exam tended to also have the lowest scores on that exam.

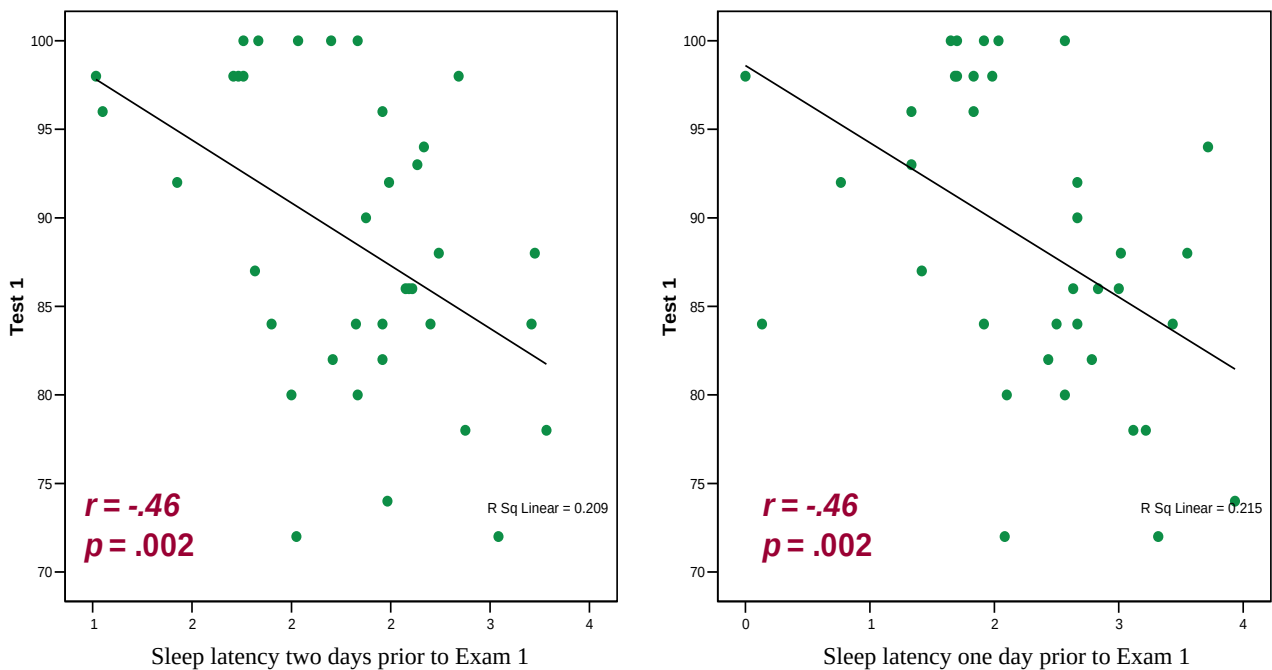


Figure 17. Scatterplots showing the significant negative correlation between exam scores and the latency to fall asleep during the two days prior to Exam 1. Note: Selected only subjects with $\geq$ five days of valid data ($n = 37$ for Exam 1).

Relationships between other actigraphic indices and performance

The Actiware program provides indices on a number of potentially important sleep variables. The relationships between these variables and Exam 1 performances are summarized briefly below:

Assumed sleep

The assumed sleep variable is determined by the Actiware scoring algorithm and reflects the amount of time between the point at which it appears that the onset of the sleep period and the time that they finally arose from sleep. This variable only counts the elapsed time from assumed sleep onset to the end of the sleep period, but does not account for brief periods of waking or sleep between those two points. This index, while only a crude estimate of sleep time, was positively correlated with Exam 1 scores for the two day and one day periods preceding the exam (two-day $r = .51, p = .001$; one-day $r = .48, p = .001$).

Actual wake time

The actual wake time variable is determined by the Actiware scoring algorithm and reflects the number of minutes of time within the assumed sleep period that were spent awake (i.e., not scored as sleep). This variable was not significantly correlated with exam performance for either the two day (two-day $r = .20, p = .12$) or one day (one-day $r = .21, p = .11$) period preceding the exam.

Sleep efficiency

The sleep efficiency variable is determined by the Actiware scoring algorithm and indicates the percent of time that a Soldier is assumed to be “in bed” or attempting to sleep that is actually spent asleep. Strong positive correlations were found with Exam scores when this variable was averaged over the two days prior to the exam (two-day $r = .60, p = .001$) as well as when evaluated only for the night preceding the exam (one-day $r = .54, p = .001$).

Number of sleep bouts

The number of sleep bouts variable is determined by the Actiware scoring algorithm and reflects the number of independent bouts of sleep identified during the assumed sleep period. The number of bouts of sleep was positively correlated with Exam 1 performance for the two-day (two-day $r = .53, p < .001$) and one-day (one-day $r = .56, p < .001$) measurement periods.

Immobile minutes

The immobile minutes variable is determined by the Actiware scoring algorithm and reflects the number of minutes within the assumed sleep period that were scored as immobile (i.e., without detectible movement). The number of minutes scored as immobile was positively correlated with exam performance when evaluated over the two days prior to the exam (two-day

$r = .58, p < .001$) as well as when evaluated for just the one day before the exam (one-day $r = .52, p = .001$).

Fragmentation index

The fragmentation index is determined by the Actiware scoring algorithm and reflects the amount of movement or disrupted sleep. Interestingly, this variable was also positively correlated with performance on Exam 1. Greater fragmentation of sleep during the two days prior to the exam was modestly associated with better exam performance (two-day $r = .29, p = .04$). This pattern was also evident when assessed only for the one day prior to the exam (one-day $r = .28, p = .05$).

Validation of SHARP

Overview

A goal of the present study was also to validate the use of the SHARP package for providing useful summary scores regarding sleep/wake and prediction of cognitive performance from actigraphy data. This process involved 1) evaluating the validity of the SHARP output variables for predicting actual militarily relevant cognitive performance (i.e., military course examination scores), 2) comparing this predictive validity to simple actigraphically scored sleep measures (i.e., without the complex mathematical modeling provided by the SAFTE Model implemented within the FAST program upon which the SHARP is built), and 3) determining the best smoothing parameters to maximize the predictive capabilities of the SHARP. From the SHARP output, average ME scores were extracted for the mission window encompassing the time of the course exam for each subject. These were then entered into correlational analyses to predict exam scores. This process was repeated for all three levels of smoothing (i.e., 0, 15, and 30 minute smoothing).

Prediction of exam scores from SHARP ME index

Without any smoothing applied to the data (i.e., 0 smoothing), SHARP ME scores were significantly correlated with Exam 1 scores ($r = .50, p = .001$), suggesting that the SHARP was effective at predicting cognitive performance. When a 15-min smoothing criterion was applied, the correlation was strengthened slightly ($r = .56, p < .001$). The best prediction was achieved, however, when the 30-min smoothing criterion was used ($r = .60, p < .001$). As evident in figure 18, it was possible to predict approximately 36 percent of the variance in exam scores based on the predicted ME scores during the mission window encompassing the exam.

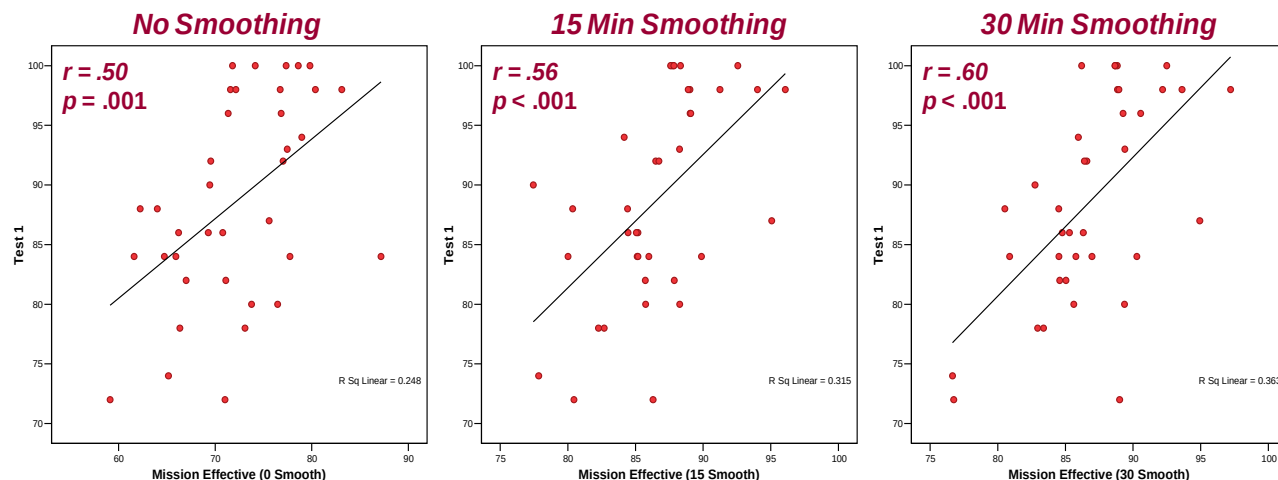


Figure 18. Scatterplots showing the significant positive correlations between Exam 1 scores and the SHARP ME score at the three smoothing levels. Note: Selected only subjects with $\geq$ five days of valid data ($n = 37$ for Exam 1).

To further evaluate the effects of the smoothing procedures on ME scores, these data were subjected to a repeated measures analysis of variance. Overall, there was a significant effect of smoothing on total ME scores, $F(2,78) = 221.33$, $p < .001$, with the no-smoothing criterion showing significantly lower ME scores ($M = 74.0$; $SD = 8.9$) than the 15-min ($M = 87.4$; $SD = 5.3$) or 30-min ($M = 87.7$; $SD = 5.3$) smoothing procedures, which did not differ significantly from one another.

Comparison of exam scores among SHARP ME color zones

The output of the SHARP classifies each subject as falling within a color-coded range (red, yellow, green) indicating the mean level of predicted effectiveness during the mission. To examine the validity of these zone categories, each Soldier was classified into these categories or zones based on the mean predicted effectiveness at during the first exam (using the 30-min smoothing criterion). Subjects were included only if they had sufficient data for model prediction (i.e., five or more days of data prior to the exam). Forty subjects met that criterion. For the first exam, 10 (25%) Soldiers fell within the green zone, 30 (75%) Soldiers fell within the yellow zone, and none of the Soldiers fell within the red zone (0%). The exam scores between the Soldiers within the Green and Yellow zones were significantly different, $t(35) = 2.06$, $p = .047$. This analysis showed that Soldiers classified as falling within the green zone (Mean Exam Score = 94.4) scored significantly higher than Soldiers within the yellow zone (Mean Exam Score = 87.3).

Prediction of exam scores from SHARP “Percent Below Criterion” index

The SHARP output provides an index that represents the percent of the mission window during which the ME score falls below a pre-specified criterion level. This criterion can be customized for the needs of the user, although the default criterion ME level is set at 90. For the present analysis, during the hour that Soldiers were taking Exam 1, the SHARP output indicated

the percent of the exam hour that was spent below an ME score of 90. Without any smoothing applied to the data (i.e., 0 smoothing), Soldiers spent an average of 95% ($SD = 22.1$) of the exam time below the criterion cutoff. However, with the application of 15-min smoothing, this value dropped to 79% ($SD = 40.2$) of time below criterion. With 30 minute smoothing, the SHARP estimated that Soldiers spent 74.9% ($SD = 42.5$) of the exam time below the criterion level. This difference among the three smoothing levels was significant, $F(2,78) = 8.84$, $p = .003$, with the no-smoothing criterion differing significantly from the 15-min and 30-min methods. To determine whether the “Percent Below Criterion index” provides any predictive validity in accounting for variability in exam scores, these index scores were correlated with exam scores. Without any smoothing applied to the data (i.e., 0 smoothing), SHARP “Percent Below Criterion” scores were completely uncorrelated with Exam 1 scores ($r = .00$, ns), suggesting that without smoothing, these index scores did not provide useful prediction of performance. As evident in figure 19, when a 15-min smoothing criterion was applied, the correlation emerged as significant and negative ($r = -.34$, $p = .02$), suggesting that Soldiers spending more of the exam period below the criterion tended to perform worse on the exam. Applying a 30-minute smoothing to the data did not significantly change this relationship ($r = -.36$, $p = .015$).

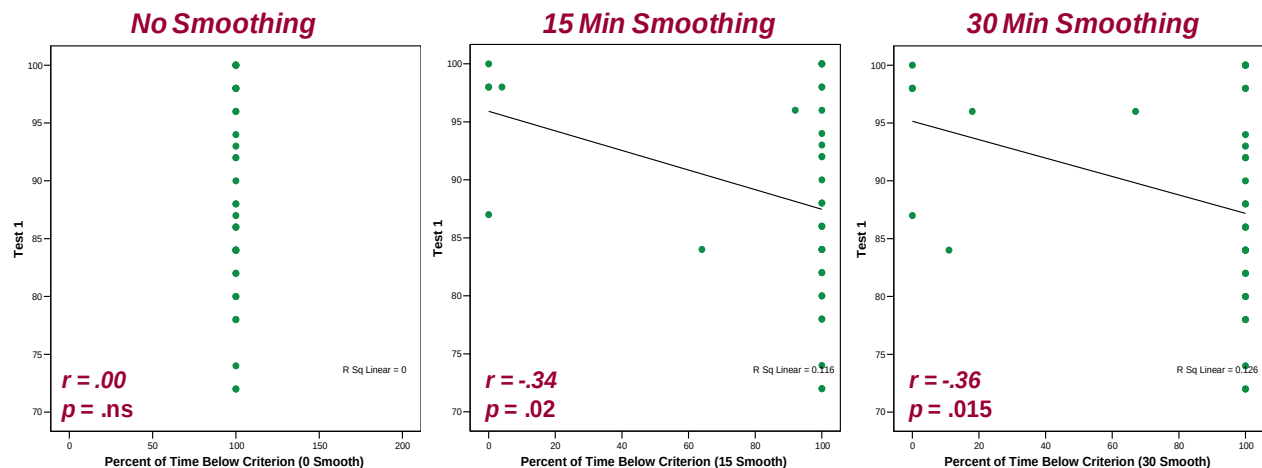


Figure 19. Scatterplots showing the correlations between Exam 1 scores and the SHARP Percent Below Criterion (ME 90) at the three smoothing levels. Note: Selected only subjects with $\geq$ five days of valid data ($n = 37$ for Exam 1).

Prediction of exam scores from SHARP Typical Effectiveness (TE) index

The SHARP output provides an index that represents the mean effectiveness score for all waking time periods that are not included in the mission window. This index was developed to provide an estimate of the average or typical level of functioning of an individual Soldier during normal waking when not engaged in the mission of interest. This Typical Effectiveness (TE) index was developed to provide a baseline estimate to allow individualized predictions tailored to each Soldier’s typical sleep-wake patterns. As with the previous analyses, data were screened to include only those subjects that had five consecutive days of actigraphic data in order to obtain a reliable estimation model. The TE index was entered into a correlation analysis to predict

exam scores for Exam 1 using the three levels of smoothing. As evident in figure 20, without any smoothing applied to the data (i.e., 0 smoothing), SHARP TE scores were significantly correlated with Exam 1 scores ($r = .45$, $p = .003$), suggesting that even without smoothing, higher TE scores were associated with better performance on the classroom exam. Subtle incremental improvement in prediction occurred with when smoothing criterion was applied at the 15-min ($r = .52$, $p = .001$) and 30-min levels ($r = .54$, $p < .001$). These findings suggest that the typical level of effectiveness based on average sleep patterns in general is related to academic performance in a military course.

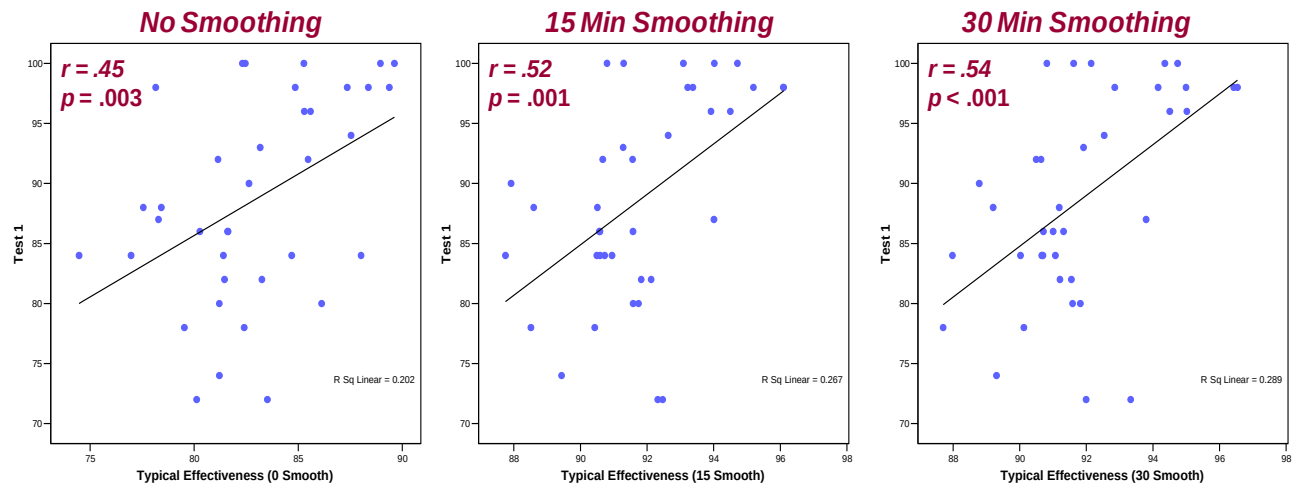


Figure 20. Scatterplots showing the significant positive correlations between Exam 1 scores and the SHARP TE scores (three smoothing levels). Note: Selected only subjects with $\geq$ five days of valid data ($n = 37$ for Exam 1).

To further evaluate the effects of the smoothing procedures on TE scores, these data were subjected to a repeated measures analysis of variance. Overall, there was a significant effect of smoothing on total TE scores, $F(2,78) = 272.27$, $p < .001$, with the no-smoothing criterion showing significantly lower TE scores ($M = 83.3$; $SD = 4.2$) than the 15-min ($M = 92.0$; $SD = 2.2$) or 30-min ($M = 92.0$; $SD = 2.3$) smoothing procedures, which did not differ significantly from one another.

SHARP Percent Change in Effectiveness (PCE) scores

Because SHARP TE scores provide an estimated baseline of typical functioning for each individual Soldier, it was hypothesized that greater prediction of performance might occur if ME scores for a particular mission window were compared to each Soldier's baseline TE score. To evaluate this hypothesis, a Percent Change in Effectiveness (PCE) score was calculated for each Soldier. This was accomplished by subtracting the Typical Effectiveness (TE) score from the ME score and dividing by the TE score (i.e., $PCE = (ME - TE) / TE$). This provides an index of the percent that an individual Soldier's effectiveness score during a mission window has deviated from that Soldier's typical level of waking effectiveness. As with the previous analyses, these were evaluated at the three levels of smoothing. On average, without smoothing, Soldiers

showed a mean PCE decline of -11.2% ($SD = 8.9$) from their typical waking level of effectiveness. With a 15-min smoothing criterion applied, this decline was reduced to -5.0% ($SD = 4.4$). The 30-min criterion was similar, with an average decline of -4.7% ($SD = 4.5$). The difference among these three smoothing approaches was statistically significant, $F(2,78) = 52.3$, $p < .001$, suggesting that the no-smoothing criterion was different from the others.

The PCE scores were entered into a correlation analysis to predict performance on the first exam. As evident in figure 21, without any smoothing applied to the data (i.e., 0 smoothing), SHARP PCE scores were significantly correlated with Exam 1 scores ($r = .41$, $p = .012$). Smoothing the data provided slight improvement in the predictive relationship, with the 15-min ($r = .46$, $p = .005$) and 30-min levels ($r = .48$, $p = .003$) all showing significant positive correlations. Overall, the average change in effectiveness from baseline that was observed during the exam period was significantly predictive of exam scores. The relationships, however, were no stronger than those observed for ME scores alone (see above), suggesting that this technique may not provide additional value in predicting performance.

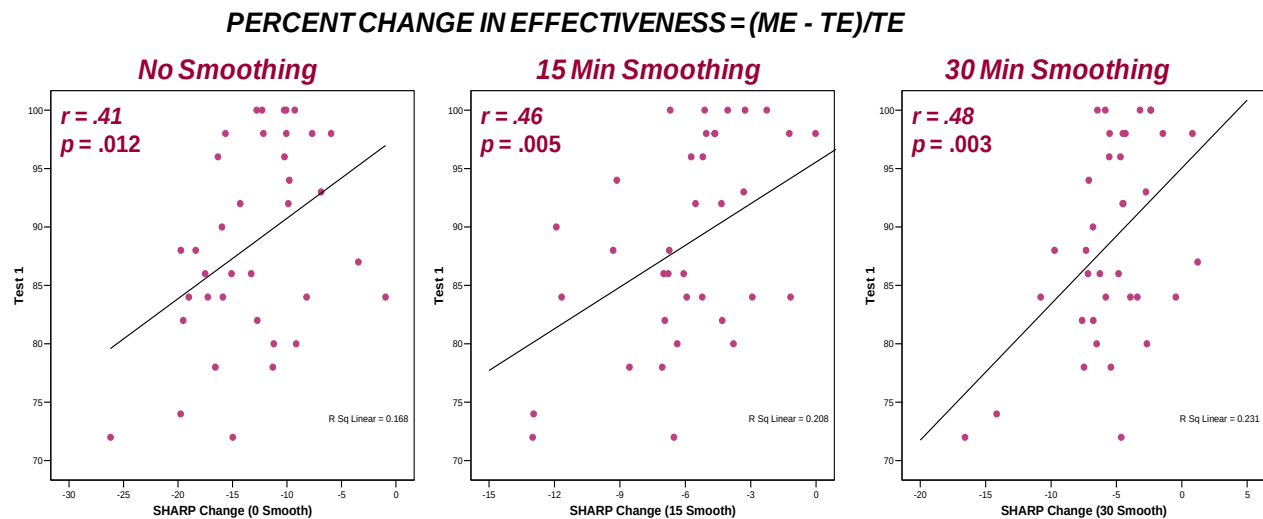
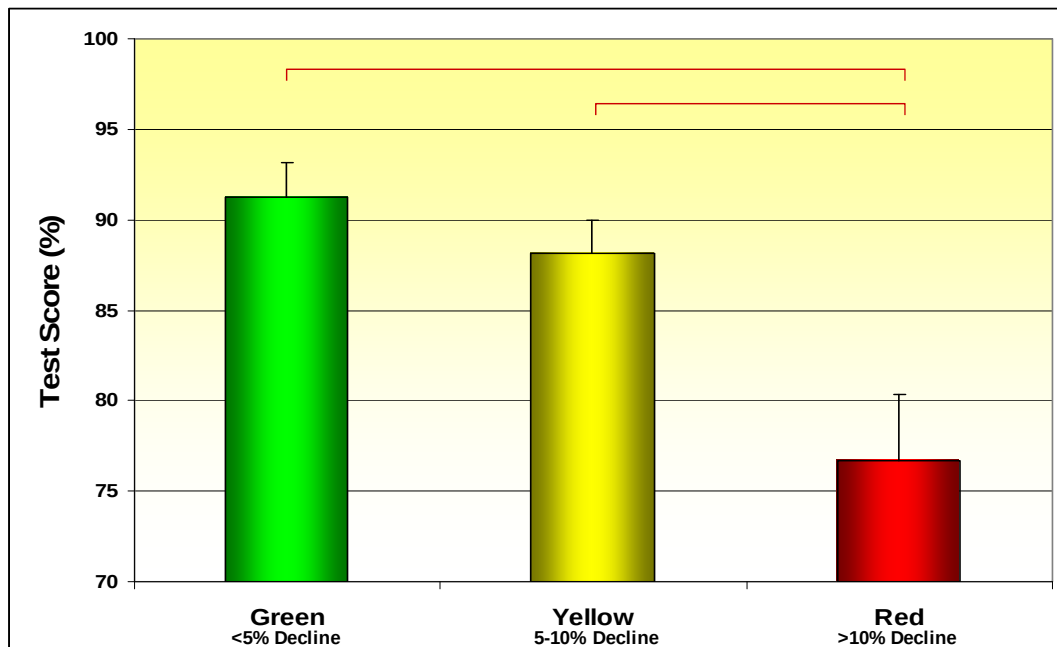


Figure 21. Scatterplots showing the significant positive correlations between Exam 1 scores and the SHARP PCE scores (three smoothing levels). Note: Selected only subjects with $\geq$ five days of valid data ($n = 37$ for Exam 1).

To further evaluate the effectiveness of this procedure at discriminating exam scores based on the severity of decline. Soldiers were assigned to one of three color coded categories based on the severity of the PCE decline (green: $< 5\%$ decline; yellow: $5-10\%$ decline; red: $> 10\%$ decline). As evident in figure 22, these three groups showed significant differences in exam scores on Exam 1, $F(2,34) = 4.54$, $p = .018$, with those identified as falling into the red category of decline showing significantly poorer exam scores relative to either the green or yellow categories.

$$\text{PERCENT CHANGE IN EFFECTIVENESS} = (ME - TE) / TE$$



ANOVA: $F(2,34) = 4.54, p = .018$

Figure 22. Exam scores on Exam 1 differed for the three groups defined on the magnitude of decline in PCE scores. Note: Selected only subjects with $\geq$ five days of valid data ($n = 37$ for Exam 1).

Prediction of exam scores from SHARP Average Sleep index

In addition to the model predictions that are provided by the SHARP, an additional column of data is also provided to summarize the average amount of sleep that Soldier has obtained across the entire period of collected data. This index is similar to the TE index in that it provides a global estimate of sleep averaged over many days rather than providing specific information targeted at the mission window. The Average Sleep index provided by the SHARP is different from the output previously described for the Actiware Sleep Analyses because the SHARP includes all periods of assumed sleep throughout the 24-hour period, including naps or other periods of inactivity that are scored as sleep. The Actiware Sleep Analyses, in contrast, only included scored sleep periods during the “in bed” period from 2100 to 1100 the next day. Thus, for some subjects, the SHARP program may overestimate sleep by also including daytime periods of inactivity in its sleep calculations. As with the analyses above, this analysis only included those subjects that had five consecutive days of actigraphic data and were analyzed for the three levels of smoothing. As expected based on these differences, SHARP Average Sleep index scores leading up to Exam 1 were only modestly correlated with total Actiware Sleep Analysis scores (i.e., 0-min smoothing: $r = .39, p = .01$; 15-min smoothing: $r = .38, p = .02$; 30-min smoothing: $r = .40, p = .01$). For the period prior to Exam 1, the Average Sleep was 7.31 hr ($SD = 2.23$) for the 0-min smoothing criterion, 7.27 hr ($SD = 2.28$) for the 15-min smoothing

criterion, and 7.23 hr ($SD = 2.27$) for the 30-min smoothing criterion, and did not differ significantly from one another, $F(2,78) = 2.11, p = .15$. These data were highly correlated with exam performance on Exam 1. As evident in figure 23, without any smoothing applied to the data (i.e., 0 smoothing), SHARP Average Sleep scores were significantly correlated with Exam 1 scores ($r = .54, p < .001$), suggesting that even with no smoothing, higher estimates of Average Sleep in the weeks leading up to the exam were associated with better exam performance. These relationships were slightly stronger with additional smoothing of the data to 15-min ($r = .60, p < .001$) and 30-min levels ($r = .64, p < .001$). The latter correlation suggests that the 30-minute smoothing estimate of the average amount of sleep obtained prior to the Exam 1 was able to account for 40% of the variance in exam scores.

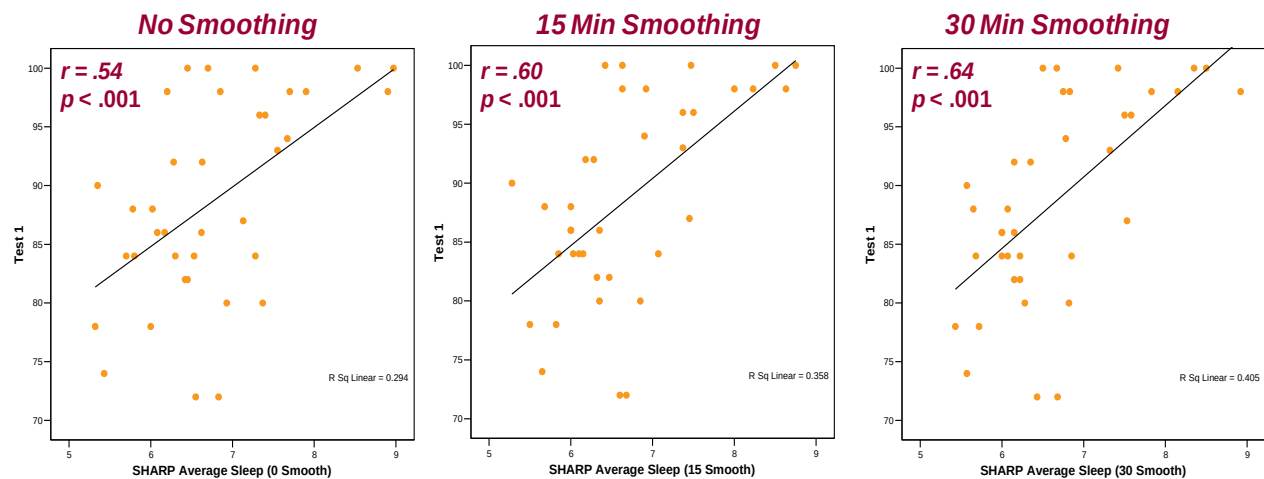


Figure 23. Scatterplots showing the significant positive correlations between Exam (Test) 1 scores and the SHARP Average Sleep scores prior to the exam (three smoothing levels). Note: Selected only subjects with $\geq$ five days of valid data ($n = 37$ for Exam 1).

All three exams combined

When data were available for all three examinations, the datasets were concatenated to form a column that included all three exams and associated columns that included the SHARP sleep scores preceding each exam. This permitted an overall evaluation of the relationships between the SHARP sleep-related variables and a single outcome variable of “exam score” across all exam administrations regardless of course or topic.

Prediction of all three exam scores from SHARP ME index. The ME scores prior to each of the three exam administrations were subjected to a repeated measures analysis of variance. Overall, there was a highly significant effect of smoothing on total ME scores, $F(2,264) = 283.01, p < .001$. On average, the no-smoothing criterion showed significantly lower ME scores ($M = 779; SD = 12.1$) than the 15-min ($M = 90.7; SD = 5.4$) or 30-min ($M = 91.0; SD = 5.3$) smoothing procedures, which did not differ significantly from one another. With no smoothing applied to the data (i.e., 0 smoothing), SHARP ME scores for the three exams combined were not significantly correlated with Exam scores ($r = .15, p = .14$). In contrast, with the application of a 15-min smoothing criterion, the correlation was significant ($r = .35, p < .001$), and was

strengthened slightly with the use of a 30-min smoothing criterion ($r = .39, p < .001$). These data suggest that with a 15- to 30-min smoothing criterion, SHARP ME scores were significantly predictive of Exam performance across courses, material, and exam times (figure 24).

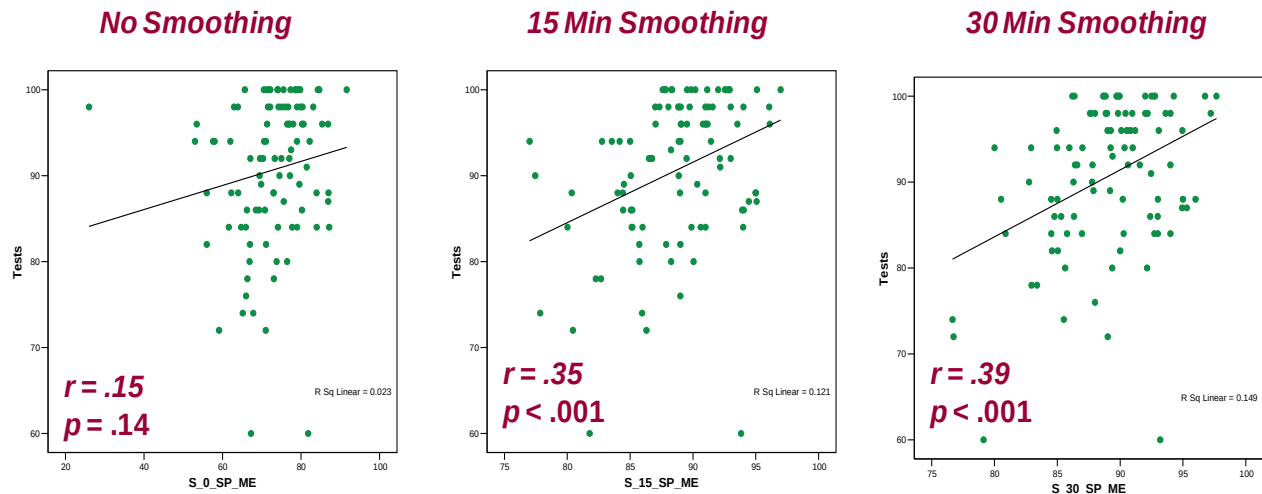


Figure 24. Scatterplots showing the significant positive correlations between all three exam scores and the SHARP ME score (three smoothing levels). Note: Selected only subjects with $\geq$ five days of valid data ($n = 97$ across three exams).

Prediction of all three exam scores from SHARP Average Sleep index. The Average Sleep as measured by the SHARP prior to each of the three exams was used to predict the exam scores. As in the previous analyses, data were only included if the subject had five consecutive days of actigraphic data. Similar to the findings reported above for Exam 1, SHARP Average Sleep index scores leading up to all three of the exams were significantly correlated with exam scores (i.e., 0-min smoothing: $r = .23, p = .03$; 15-min smoothing: $r = .25, p = .012$; 30-minute Smoothing: $r = .27, p = .008$). These relationships are illustrated in figure 25. The SHARP estimates of Average Sleep were highly correlated with total sleep time for the two days prior to each exam as measured by the Actiware Sleep Analysis scoring system for the “in bed” period from 2100 to 1100 (i.e., 0-min smoothing: $r = .54, p < .001$; 15-min smoothing: $r = .55, p < .001$; 30-min smoothing: $r = .57, p < .001$).

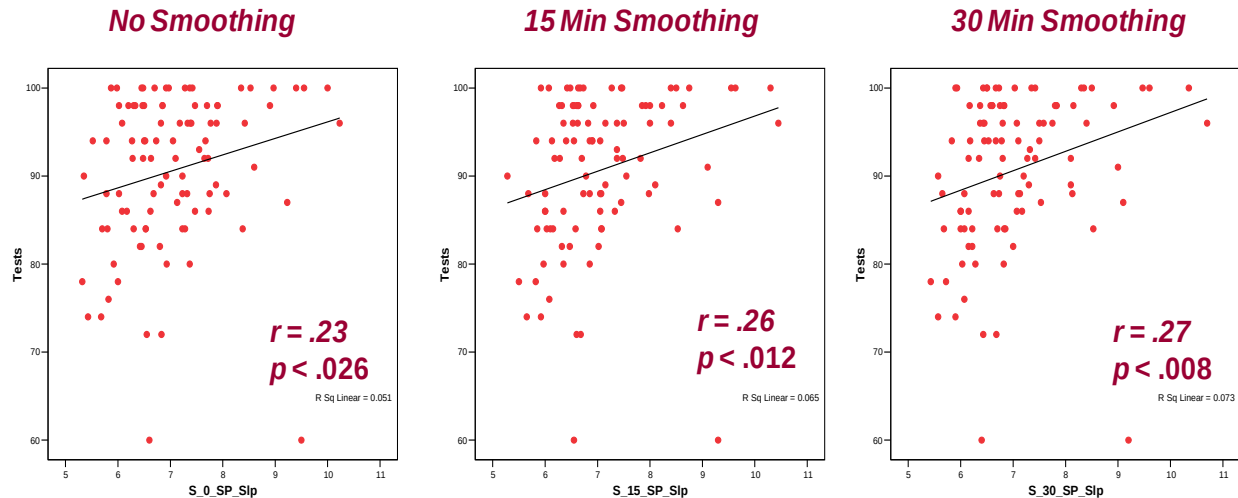


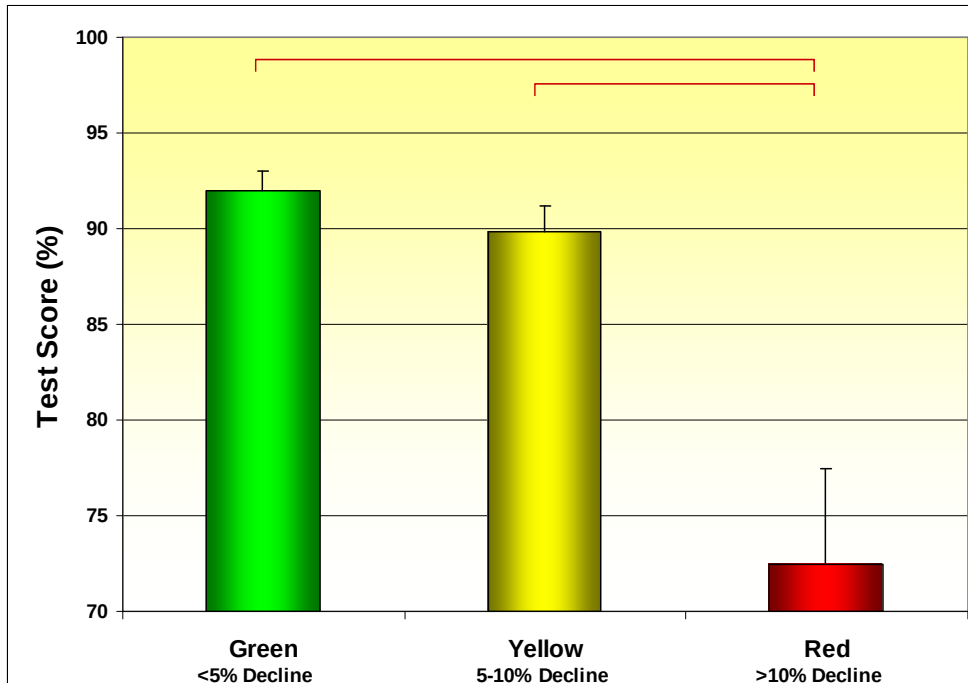
Figure 25. Scatterplots showing the significant positive correlations between all three exam scores and the SHARP Average Sleep scores prior to each exam (three smoothing levels). Note: Selected only subjects with $\geq$ five days of valid data ($n = 97$ across three exams).

SHARP PCE scores for all three exams. As described previously, ME scores were compared to a baseline level of performance estimated from TE scores. The usefulness of the PCE score was evaluated across all three exam administrations. As with the previous analyses, these were evaluated at the three levels of smoothing. On average for the three exams, without smoothing, Soldiers showed a mean PCE decline of -4.9 ($SD = 14.5$) from their typical waking level of effectiveness. At a 15-min smoothing criterion, this decline was reduced to -1.4% ($SD = 6.0$). The 30-min criterion was similar, with an average decline of -1.3% ($SD = 5.9$). The difference among these three smoothing approaches was statistically significant, $F(2,228) = 13.72$, $p < .001$, suggesting again that the no-smoothing criterion was different from the others.

As with the previous analysis for Exam 1, the PCE scores were entered into a correlation analysis to predict performance on all three exams. Without any smoothing applied to the data (i.e., 0 smoothing), SHARP PCE scores showed only a marginally significant correlation with the three exam scores ($r = .19$, $p = .06$). The 15-min smoothing resulted in a stronger relationship ($r = .32$, $p = .002$) and the 30-minutes smoothing criterion provided slightly better prediction ($r = .36$, $p < .001$).

Finally, the effectiveness of the three color-coded change categories at discriminating exam scores was evaluated. In the same manner as described previously for Exam 1, Soldiers were assigned to one of three color coded categories based on the severity of the PCE decline (green: $< 5\%$ decline; yellow: $5\text{--}10\%$ decline; red: $> 10\%$ decline). Figure 26 shows that these three groups were significantly different with regard to exam performance, $F(2,934) = 11.8$, $p < .001$, with those identified as falling into the red category of decline showing significantly poorer exam scores relative to either the green or yellow categories.

$$\text{PERCENT CHANGE IN EFFECTIVENESS} = (ME - TE) / TE$$



ANOVA: $F(2,94) = 11.8, p < .001$

Figure 26. Color-coded categories defined on the magnitude of decline in Percent Change in Effectiveness scores across all three examinations combined. Note: Selected only subjects with $\geq$ five days of valid data ($n = 97$ for three exams).

Discussion

The primary objectives of this study were to obtain actigraphic measurements of sleep in a sample of Soldiers during military education/training and to use those data to examine the relationship between sleep and militarily relevant performance and to validate previously established models of sleep and cognitive performance. These objectives were accomplished. The present study yielded several important outcomes:

1. Actigraphic measurement of sleep can be obtained successfully and unobtrusively within a militarily relevant environment (worn by Soldiers whose training included limited field conditions and vigorous physical training).
2. Actigraphic data can be reliably organized and scored using automated algorithms to identify poor quality data.
3. Semi-automated Actiware sleep scoring revealed that most Soldiers are presently obtaining sub-optimal amounts of sleep per night during military education and the amount of measured sleep obtained by these Soldiers is directly and significantly related to military course performance.

4. The automated SHARP program was effective at predicting military course performance based on the implementation of mathematical prediction models that incorporated the amount of recent sleep and time of day of the exam.

Together, these findings provide preliminary support for the use of actigraphy to measure sleep/rest in militarily relevant settings. These results also demonstrate the importance of adequate sleep for course learning and retrieval, and suggest that automated methods such as the SHARP provide valid and militarily relevant prediction of performance.

While actigraphy has been used successfully for several decades within laboratory and some field settings, there have often been difficulties with subject compliance and the ability to obtain reliable and valid measurements. As the ultimate goal is to use actigraphy in actual combat operations to aid in sustaining the health of the force and to optimize command decisions regarding manpower allocation, it is critical to demonstrate that actigraphy can be obtained reliably and unobtrusively with Soldiers in militarily relevant settings. The present data provide a mixed picture regarding the ability to obtain useful data in a military setting. Following extensive data cleaning and validity checks, 59% of the sample had reliable and valid data that were complete enough for inclusion in the final statistical analyses. Of those data that were obtained, the results and predictive validity were impressive. The correlations between the amount of obtained sleep and subsequent course performance were highly significant. Unfortunately, the present study suffered from a high level of data loss. Twenty-three percent of the returned Actiwatchs failed to record any activity data. From the post-collection analysis, it was not clear whether this loss of data was due to a mechanical malfunction, battery failure, or human error in initializing the actigraphs at the start of the study. Thus, the data that were collected were highly predictive and meaningful, but a major limitation to this method that must be addressed by future studies is the unacceptably high failure rate of the actigraph recording devices.

One of the major difficulties in using actigraphy in a combat or other field environment is the extent of pre-processing and data management that is necessary in order to transform the large data streams into a form that is readily usable for describing current sleep levels and making performance predictions. The present effort attempted to outline the elementary steps necessary to take raw activity counts from the Actiwatch and transform them into a form that could be analyzed in standard statistical packages, spreadsheets, or entry into statistical models. Simple data quality assessment tools were created and automated in Microsoft Excel to evaluate the validity of the obtained data. This process was accomplished via simple macros written in Microsoft Excel and which could potentially be implemented into self-contained free-standing software programs. These processes were then compared to manual scoring/assessment procedures and were found to produce agreement exceeding 98% between the manual and automated validation procedures. These findings suggest that it may be possible to automate many of the pre-processing and validation stages to make actigraphic analysis of sleep data feasible and expedient in field environments.

In order for actigraphic measurement of sleep to be useful in military settings, it must be shown to be meaningfully related to relevant indices of performance. The present results showed that Soldiers undergoing these military training courses obtained an average of 5.8 hr of

actigraphically measured sleep per night, an amount that is significantly less than the 7 to 9 hr that is generally recommended as healthy for adults (<http://www.sleepfoundation.org>). In fact, for the sample as a whole, greater than 95% of the participating Soldiers obtained less than the minimum recommended 7 hr of sleep per night on average throughout the duration of the course. This suggests that the sleep habits of Soldiers while attending these courses is generally below what is considered adequate for long term health of normal adults. Poor sleep is related to a variety of health problems and chronic illnesses. The potential consequences of prolonged inadequate sleep in terms of physical health, well-being, and readiness of the force are issues that should be considered in the design and implementation of future military training programs.

It was also of interest to determine whether the amount of obtained sleep during military training was meaningfully related to course performance. Accordingly, outcome data were collected in the form of exam scores from the various courses. Due to small class sizes and availability of specific Soldier samples, the groups comprised several different military courses, including the Warrant Officer Candidate School and several different tracks of the Non-Commissioned Officer Academies. The outcome measures included performance on the first exam and performance on all course exams. It is important to note that these exams differed in content across the various courses. Despite the variability in exam content, overall percent correct on pooled exam scores across all courses was significantly predicted by several sleep related variables. In some cases, the average amount of sleep obtained by Soldiers accounted for approximately 40% of the variance in exam scores. This finding underscores the importance of adequate sleep on cognitive functioning, including learning and retention of military relevant information. Soldiers that obtained more sleep were generally more able to benefit from training, retained more information, and obtained higher exam scores than their peers that obtained less sleep. Also, as evident in figure 23, the amount of sleep obtained was related to the variability or “instability” of exam scores. Close examination of figure 23 shows that those Soldiers that consistently averaged the highest amounts of sleep obtained consistently high exam scores, whereas those that averaged low levels of sleep obtained inconsistent performances on the exams, with some doing quite well and others receiving failing or only marginally passing grades. Thus, one of the most important factors associated with performing well in these military courses was the degree to which a Soldier obtained adequate sleep on a consistent basis.

One of the major objectives of this study was to determine whether the automated SHARP program could be used to download, score, and predict militarily relevant performance. On the whole, this objective was met. The SHARP was effective at initializing and downloading actigraphic data. Although some technical difficulties emerged in the process of running large groups of data with the program, the program successfully ran the sleep and performance models on all available data and provided the required output as designed. The resulting output of the SHARP was found to be highly predictive of Army course performance. These preliminary results suggest that the predictive model and output from the SHARP are equal to or better than raw actigraphy sleep scores alone, and are more easily obtained and organized than other available methods of sleep scoring. Furthermore, the present findings suggest that at least some smoothing of the data is necessary to obtain adequate predictive models, and that of the three smoothing levels tested, the 30-min time smoothing criterion generally provides the best predictive models. It is, therefore, recommended that the 30-min smoothing criterion be used as the standard default setting for most purposes until further data are available.

The SHARP provides a single user interface to streamline the process of downloading actigraphs and automate the process of using FAST™ to analyze data. In addition, the SHARP output provides an easy to interpret summary of the sleep/wake history and predicted cognitive performance of unit personnel. The output of the SHARP is designed to be mission relevant and present only those data that are likely to be critical to the commander's decision-making process. The present findings suggest that the SHARP can be used to download, organize, and make rapid decisions regarding the overall readiness of a unit or of specific personnel for tasks requiring cognitive performance.

Limitations

The present study suffers from a number of weaknesses. First, the significant loss of data was problematic and suggests that the use of actigraphy in non-laboratory studies is still prone to considerable mechanical and/or human error. Future studies will need to be vigilant about minimizing these problems either through automation or more effective staff training. Second, the loss of data clearly limits the generalizability and power of the study. Future studies with larger samples will be necessary to establish the reliability of these findings. Third, the outcome variable in the present study was "exam score" without regard for the specific content of the course, the teaching methods used, or other variables that likely added considerable noise to the data. On one hand, the fact that actigraphic sleep was able to predict performance on such a loosely defined outcome variable as exams to the robustness of sleep as a factor in cognitive performance, the study would have been more precise had the dependent variable been the same across all subjects. Furthermore, the present findings are only applicable to the types of course content presented in the Warrant Officer Candidate School and NCO academies. These dependent variables represent general learning and retention of textbook military content. While the present findings suggest that performance on this type of material in an academic environment is adversely affected by inadequate sleep, it remains unclear whether actigraphic measurement of sleep is related to other types of cognitive performance in the field. These are issues to be examined in subsequent research.

Conclusions

This study was conducted to determine if performance models based on sleep history and activity levels could predict academic exam performance in the NCO Academy and Warrant Officer Candidate School at Fort Rucker. Analysis of the sleep and activity data from six classes showed that independent of the Soldiers' academic background, type of school, or exam content, the amount of sleep during the two nights previous to an academic exam accounted for up to 40% of the variance in exam scores. Prior activity levels were not predictive. These data support the use of actigraphy for forecasting cognitive performance in military personnel. There was also a significant effect on changes in Soldier effectiveness predicted by models based on these data. These data suggest that actigraphy can be easily and unobtrusively collected from military personnel and provide useful real-time information about the readiness of Soldiers and aircrew. Such easily obtained data may improve military decision-making as well as potentially save lives and equipment.

References

- Army Training Requirements and Resources System. 3 Jun 09. ATRRS Course Catalog. <https://www.atrrs.army.mil/atrrscc/>.
- Actiwatch®16 / Actiwatch®64 / Actiwatch®-L / Actiwatch®-Score Instructional Manual. 2001. Bend, OR: Mini Mitter Company, Incorporated.
- Balkin, T., Thome, D., Sing, H., Thomas, M., Redmond, D., Wesensten, N., Williams, J., Hall, S., and Belenky, G. 2000. Effects of sleep schedules on commercial motor vehicle driver performance. Washington: Department of Transportation Federal Motor Carrier Safety Administration. Report No. DOT-MC-00-133.
- Belenky, G., Wesensten, N.J., Thorne, D.R., Thomas, M.L., Sing, H.C, Redmond, D.P., et al. 2003. Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study. Journal of Sleep Research, 12: 1-12.
- Caldwell, J.A., Smythe, N.K., Caldwell, J.L., Hall, K.K., Norman, D.N., Prazinko, B.F., Estrada, A., Johnson, P.A., Crowley, J.S., and Brock, M.E. 1999. The effects of Modafinil on aviator performance during 40 hours of continuous wakefulness: a simulator study. Fort Rucker, AL: United States Army Aeromedical Research Laboratory. Report No. 99-17.
- Caldwell, J.L. and Caldwell, J.A. 1993. A comparison of sleep scored from electroencephalography to sleep scored by wrist actigraphy. Fort Rucker, AL: United States Army Aeromedical Research Laboratory. Report No. 93-32: 3.
- Carskadon, M.A. and Dement, W.C. 1989. Normal human sleep: An overview. In M.H. Kryer, T. Roth, and W.C. Dement (Eds.), Principals and practice of sleep medicine. Philadelphia: W.B. Saunders Company.
- Cole, R.J. and Kripke, D. 1988. Progress in automatic sleep/wake scoring by wrist actigraph. Sleep Research, 17: 331.
- Comperatore, C.A., Chiaramonte, J.A., Pearson, J.Y., Stone, L.W., Hess, G.L., and Boley, K.O. 1993. Activity/rest patterns of instructor and rated student pilots during rapid transitions from daytime to nighttime duty hours at Eastern Army Aviation Training Site. Fort Rucker, AL: United States Army Aeromedical Research Laboratory. Report No. 93-16: 7.
- Comperatore, C.A., Lieberman, H.R., Kirby, A.W., Adams, B., Crowley, J.S. 1996. Melatonin efficacy in aviation missions requiring rapid deployment and night operations. Aviation, Space, and Environmental Medicine, 67: 520-524.

- Dinges, D., Pack, F., Williams, K., Gillen, K., Powell, J., Ott, G., Aptowicz, C., and Pack, A. 1997. Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. Sleep, 20(4): 267-277.
- Hursh, S. R., Redmond, D. P., Johnson, M. L., Thorne, D. R., Belenky, G., Balkin, T. J., et al. (2004). Fatigue models for applied research in warfighting. Aviation, Space, and Environmental Medicine, 75(3 Suppl), A44-53.
- Kripke, D., Webster, J.B., Mullaney, D.J., Messin, S., and Mason, W. 1981. Measuring sleep by wrist actigraph. La Jolla, CA: University of California. Report No. 3.
- Krueger, G.P. 1989. Sustaining military performance in continuous operations: Combatant fatigue, rest, and sleep needs. In Gal, R., and Mangelsdorff, A.D. (Eds.), Handbook of military psychology, 255-277. New York: John Wiley and Sons.
- LeDuc, P. A., Riley, D., Hoffman, S. M., Brock, M. E., Norman, D., Johnson, P. A., Williamson, R., and Estrada, A. 2000. The effects of sleep deprivation on spatial disorientation. U.S. Army Aeromedical Research Laboratory Technical Report, No. 2000-09.
- Lieberman, H.R., Bathalon, G.P., Falco, C.M., Morgan, C.A., Niro, P.J., and Tharion, W.J. 2005. The fog of war: Decrements in cognitive performance and mood associated with combat-like stress. Aviation, Space, and Environmental Medicine, 76(7), II, C7.
- Lieberman, H.R., and Coffey, B.P. 2000. Preliminary finding from a new device for monitoring performance and environmental factors in the field. In K. Friedl, H. Lieberman, D.H. Ryan, and G.A. Bray (Eds.), Pennington Center Nutrition Series: Vol. 10 Countermeasures for Battlefield Stressors. 126-157. Baton Rouge: Louisiana State University Press.
- Miller, D.B. 2005. Sleep and predicted cognitive performance of new cadets during cadet basic training at the United States Military Academy. Monterey, CA: Naval Postgraduate School. xv.
- National Sleep Foundation. 14 Sep 07. 2006 Sleep In America Poll. <http://www.sleepfoundation.org>.
- Sadeh, A., Alster, J., Urbach, D., and Lavie, P. 1989. Actigraphically based automatic bedtime sleep-wake scoring; validity and clinical applications. Journal of Ambulatory Monitor, 2: 209-216.
- U.S. Army Warrant Officer Career College. 28 May 09. Candidate Course. <http://usawocc.army.mil/WOES/wocs.htm>.
- Webster, J.B., Kripke, D.F., Messin, S., Mullaney, D.J., and Wyborney, G. 1982. An activity-based sleep monitor system for ambulatory use. Sleep, 5: 389-399.

Wertz, A.T., Wright, K.P., Ronda, J.M., and Czeisler, C.A. 2006. Effects of sleep inertia on cognition. Journal of the American Medical Association, 295(2): 163-164.

Manufacturer's list

Actiwatch® and Actireader
Mini Mitter Company, Incorporated
20300 Empire Avenue, Building B-3
Bend, OR 97701 USA
(800) 685-2999
(541) 322-7272
(541) 322-7277 (FAX)
mm@minimitter.com
www.minimitter.com



DEPARTMENT OF THE ARMY
**U.S. Army Aeromedical
Research Laboratory**
Fort Rucker, Alabama 36362-0577