Final Report:
Measuring Human Fatigue With the BLT Prototype

Date:	9.24.09
Project Title:	Measuring Human Fatigue with the BLT Prototype
Grantor:	National Institute for Occupational Safety and Health (NIOSH) http://www.cdc.gov/niosh/
Grant Number:	5R44OH007664-03
Principal Investigator:	Dr. Theodore D. Langley, Ph.D.
Project Title:	Measuring Human Fatigue with the BLT Prototype
Institution:	Bowles-Langley Technology, Inc.
Contact Information:	Bowles-Langley Technology, Inc. Suite 200 1801 Clement Avenue Alameda, CA 94501 Phone: 510 864 3111 Fax: 510 864 3112 tlangley@bowles-langley.com hbowles@bowles-langley.com
Co-Investigators:	Dr. Anneke Heitmann, Ph.D. Dr. Deborah L. Schnipke, Ph.D. Dr. J. Wesson Ashford, MD. Ph.D. Karen Hansen, MD. Henry M. Bowles (Project Administrator)
Project period:	06/01/2002 - 07/31/2009

Table of contents:

Abstract

During this project investigators conducted three different experiments to validate an alertness testing system developed by Bowles-Langley Technology "BLT". The system is designed to screen workers prior to work to ensure they are not impaired by fatigue or by other factors. Intended users include truck drivers, power plant operators, pilots and others engaged in high-risk activities.

Benefits to society include reducing accidents from fatigue or other impairment sources (illness, alcohol, drugs, etc.) and reducing operating and insurance costs. It is estimated that 70% of accidents involve human error with fatigue or impairment as a root or contributory cause. Current industry practice deals with this problem by managing shift work scheduling, educating operators and operator self-monitoring. However, there are no current minimal standards for measuring fatigue and impairment that are practical for actual workplace use. The test developed during this project is a valuable new tool for assessing fatigue in the workplace.

To be practical in the workplace a fitness for work test must be short so as not to significantly delay the start of work. A two-minute test is acceptable but a 5 - 10 minute test would not be. In addition, the test should not depend on language skill and should not be excessively difficult. At the same time it must also be valid and reliable in detecting impairment. This project tested whether BLT Alertness Test meets these criteria.

The BLT test displays graphic shapes on a checkerboard background. Subjects must determine if all the shapes are the same or if one shape is different. Test time for a 50-item test is about 2 minutes. The scoring algorithm takes speed, accuracy and item difficulty into account and then compares the computed score to a previously established individual baseline score. Thus, each subject's performance is measured against his or her personal baseline rather than in comparison to others. The test incorporates both minimal standards and a permitted baseline deviation both of which can be adjusted as required.

Data privacy is managed through PIN numbers and code names with separate passwords for the system administrator and local managers. Each worker/operator is assigned a PIN.

The project consisted of three trials intended to achieve three critical aims. The first aim was to determine the relative difficulty of the different possible shape combinations with a goal of improved test reliability. The second aim was to measure validity (sensitivity to fatigue induced impairment) and reliability (test-retest stability). The third aim was to assess the acceptability and feasibility of the system for managers and operators in an operational setting. All three aims were successfully attained.

Results indicate that testing for fatigue and impairment in the workplace is both feasible and practical.

Section One

Highlights and Significant Findings

The BLT Alertness test was shown to be a useable and understandable short test with good operator acceptance. The protocols for Aims 1, 2 and 3 included testing 155 individuals from a variety of educational backgrounds, ages and ethnicities. All subjects were able to understand the test and respond properly using the left and right arrow keys on a computer keyboard. Very few subjects indicated that the test was too fast, too difficult or too long. This finding confirms the premise that the BLT Alertness Test is acceptable for the general working population. It remains to be seen if the proper balance has been established between test difficulty and test sensitivity and reliability. Certainly some of the derivations currently under review, including a spin version with rotating shapes, would be more difficult and challenging.

For the Aim 1 stability trial the project effectively accomplished the goal of measuring the different difficulty levels or IRT (Item Response Theory) difficulty for each of the shape combinations. Each combination was then assigned a level of difficulty number that was incorporated into the scoring algorithm in subsequent testing.

The Aim 2 Validation Trial experiment found that the BLT Test is able to detect severe fatigue impairment and that results generally track circadian cycles. Further refinements of test design and scoring algorithms show potential to establish the BLT Test as a standard fatigue test. Used by managers with a clear understanding of the limits of current test reliability there could be beneficial results in many work environments.

The Aim 3 Implementation Trial was designed to assess the feasibility of the BLT Test in an operational setting. This trial, conducted at the University of Maryland Teaching Hospital with 20 working medical doctor interns, found that computerized impairment testing was acceptable to users in a busy workplace. Survey data showed that the process increased awareness of alertness and performance on the job. Most of the doctors felt it was not an invasion of their privacy and testing was not considered intrusive considering the potential benefits. Such findings suggest that impairment testing with computers both before and during shift work is feasible and would be viewed positively by a majority of workers.

In addition, the test software, running entirely on the Internet, proved to be robust and reliable. Investigators from locations across the U.S. were able to track results as the data were generated throughout the experiment. In a commercial installation unit managers will be able to monitor the condition of the workforce in real time from any location.

Translation of Findings

A short, non-linguistic test that can effectively measure human fatigue impairment, as well as other types of impairment, is the *Holy Grail* of fitness-for-work testing. This study is a significant step toward this goal in that it demonstrated the potential of a short test. Provided the developed software can meet standards of validity, reliability and specificity there are no serious technical impediments to implementation in selected industries. But any change in workplace management that might affect work hours or human resource requirements must be carefully

balanced against potential benefits. The best approach might be introduction on a voluntary basis. Particular attention has been paid to keeping user experience simple and transparent for this reason.

As this and similar tests are more thoroughly developed and tested their implementation will have positive effects on safety in the workplace as the result of two factors: first, the actual screening-out of impaired operators and, second, the deterrent effect on behavior of having such a system on site. This potential outcome is expected to result in a reduction in accidents caused by operator impairment in industries reliant on individual operator vigilance and executive decision efficiency. Industries include aviation, highway transportation, maritime transportation, heavy construction, etc. An integrated system tracking operator circadian and operator scheduling data with alertness testing as required is the next step towards a significant breakthrough in this field.

The results of this study should encourage further implementation trials at worksites worldwide. A statistical analysis of before and after accident rates over a one or two-year period will provide data for cost analysis based comparisons. A measurable reduction in accidents and errors can be then quantified with before and after operating costs, medical insurance costs and liability insurance costs, plus non-quantifiable human costs in pain and suffering. This assumed saving could then be compared with the implementation, operating and administrative costs of the alertness screening system to determine economic efficiency and practicality. Supporting evidence of this kind will encourage regulators and industry leaders to reexamine current practice to include fitness for work testing as a to compliment circadian shift work scheduling, worker training and other measures for ensuring alert operators.

Outcomes and Impacts

Current procedures in hazardous and dangerous operations do not truly protect the lives of the thousands of individuals who put their trust in the hands of operators on the assumption that they are alert. Trials such as this one show that practical methods that do a better job of screening are available.

Mandatory drug and alcohol testing may reduce drug use but this activity needs to be re-examined in view of its actual impact on accidents caused by impairment. A new policy offering companies the option of fitness for work screening in conjunction with drug testing and/or as a substitute for drug testing should be explored.

The study showed that a shape recognition task has considerable sensitivity to fatigue and may be sufficient to screen for severely fatigued workers. Human fatigue is generally considered to be difficult to quantify and measure so, in theory, a test that is sensitive to fatigue will also be sensitive to other causes of impairment including drug and alcohol effects, but further testing with alcohol in particular will need to be done to confirm this hypothesis.

Bowles-Langley Technology, Inc. is actively pursuing other test designs and is continuing to improve the current test based on the finding of this experiment.

List of Terms and Abbreviations

BLT Test - Standard test used by Bowles-Langley Technology, Inc. to measure human impairment.

Dwell Time – Time in seconds a screen item is shown before the next screen in the BLT test.

IRT - Item Response Theory - A statistical method of scoring that takes response time and the difficulty of different test items into account.

IRT Difficulty - A number assigned to each item combination taking accuracy and response time into consideration.

Karolinska Sleepiness Scale - A Visual Analog Scale for measuring subjective feelings of sleepiness.

OTMI (Operational Temporary Mental Impairment) A temporary condition of significantly decreased mental efficiency. Condition is characterized by slow and incorrect response to stimuli and reduced executive function. This condition may be caused by fatigue but could also be the consequence of alcohol use or drug use.

PVT - Psychomotor Vigilance Task - The PVT is a well-validated 10-minute laboratory test of behavioral alertness that is widely used to obtain an estimate of performance limits in alert and drowsy subjects, developed by D.F. Dinges and colleagues at the Walter Reed Medical Hospital in Washington, DC. Test is presented with a hand-held device using flashing lights.

SRT - Simple Reaction Time - a measure of reaction time only.

Thayer Activation Scale - A Visual Analog Scale for measuring subjective feelings of sleepiness.

VAS - Visual Analog Scale - A scale that attempts to measure a subjective feeling by the use of a horizontal line with extreme feelings indicated at each end of the line. For example 0 = very fatigued, 10 = not fatigued at all.

Section Two – Scientific Report

Background

Bowles-Langley Technology, Inc. entered the alertness testing field in 1998 with the construction of a series of proof-of-concept prototypes. The initial production units were dedicated testers using individual smart cards to identify workers and to store test baselines. The devices were installed in a number of commercial settings to establish reliability and user acceptance. Worker acceptance was found to be excellent and scoring results showed sufficient correlation with fatigue impairment to justify further experimentation. A Phase I NIOSH SBIR grant partially funded this process.

Photo 1, Original BLT tester as used in preliminary studies. Unit stores subject data on smartcards.

In 2000 BLT began experiments with running the testing software on desktop servers and on the Internet rather than on individual dedicated testers. The Internet with a high-speed connection was found to be a reliable platform and suitable for clinical trials with matched laptop computers.

Description of the BLT Alertness Test

The BLT Alertness Test is a shape recognition task that is easy to learn. The design does not simulate any particular job function as, for example, a driving simulation for truck drivers, but challenges a number of key brain functions that are necessary for all jobs. Specifically, the test measures reaction time, decision-making speed, orientation and hand-eye coordination. The software provides for a learning period (typically 10 tests) for each individual to achieve a stable baseline level of performance. This baseline is then stored by the system. In the workplace a worker's test results are compared to his/her personal baseline. To ensure data privacy each worker is assigned a PIN number and password for signing into the system.

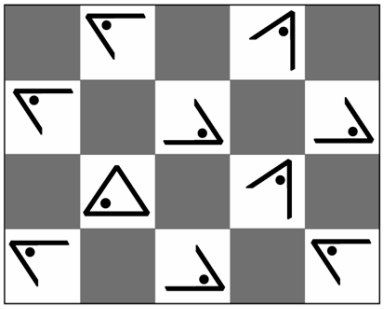

Fig. 1, Sample Screen from the BLT Alertness test. In this case one item is different so the correct response would be to press the NO key (left arrow).

The administrative account is password protection and access can be limited to a single administrator.

Initial Studies and Aims

BLT conducted a number of preliminary studies using the original BLT Tester prototype. A clinical trial with a one-night sleep deprivation period was conducted at the University of San Francisco in 1998. This was followed by second clinical trial, under Phase 1 of the current SBIR grant, at the Circadian Technology Sleep Lab demonstrated positive correlations with circadian factors. This study involved 10 subjects undergoing a one-night sleep deprivation period. Subjects were tested with the BLT Test and with a variety of test measures including the PVT Test, driving simulator and EEG measures and subjective questionnaires including the Thayer Sleepiness Scale.

These trials indicated a need for further test refinements to improve stability, sensitivity and specificity. Investigators also recommended a feasibility trial to assess the technology in a working environment. This laid the groundwork for the Phase II experiments.

Specific Aims of the Phase II Experiment

The project had had 3 specific aims:

Aim 1. Stability Trial: To improve the reliability/stability of the software by modifying the graphics used, changing the way they are presented, and changing the way the test is scored.

Aim 2. Validation Trial: To test the validity and reliability of the improved software and scoring algorithm in a sleep-deprivation trial.

Aim 3. Implementation Trial: To assess the feasibility of fitness-for-duty testing in an operational setting.

Aim 1 - Completion Report

Aim 1 Background

The alertness test draws on a library of 100 graphic shapes that are grouped in 25 families or sets of 4 shapes each (Fig. 2). The shapes in each family are similar but slightly different. Subjects are presented with a screen with all the shapes the same or, alternatively, one shape is different but is from the same family of 4 similar shapes. Speed and accuracy in detecting when one shape is different is the key ability measured by the test.

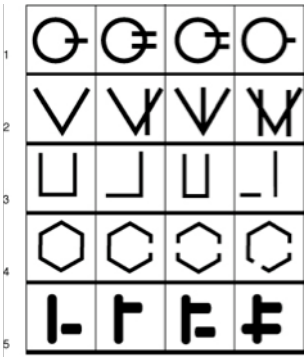

Fig. 2, Shape families 1 – 5.

It was recognized in the data and confirmed by subject reports that some shape combinations were more difficult than others. The first version of the test selected from the shape library randomly. As a result some tests could be composed of shape combinations that were more difficult than others.

To improve test-to-test stability a system had to be developed to insure all tests are of approximately equal difficulty. At the same time, the investigators needed to retain some element of random shape presentation to prevent subject memorization. The first step in this process was to measure the difficulty of each possible shape combination to establish a scale of relative difficulty for all combinations. This was the basis of the Aim 1 experiment. Special software was written designed to present all the shape combinations through a series of ten 15 - 30 minute sessions and to record speed and accuracy of subject responses for each shape combination. With this information about each shape combination the test algorithm could then take item difficulty into account. More "credit" could be given for difficult items. In addition, the item selection algorithm could be made to select items so that every test has the same number of difficult and easy items.

The Aim 1 protocol called for recruiting at least 100 subjects (115 were ultimately recruited) and for each subject to complete all 10 sessions.

Aim 1 Procedures

Each set of 4 similar shapes in the shape library may be presented in 16 different combinations.

AA	AB	AC	AD
BA	BB	BC	BD
CA	CB	CC	CD
DA	DB	DC	DD

Fig. 3, there are 16 possible shape combinations for the 4 similar shapes comprising a shape familiy. The program creates a YES item by selecting 10 identical shapes or, alternatively, creates a NO item by selecting 9 identical and one different shape. Note that AB is not the same as BA because AB assumes A is the dominant shape with one B, and BA assumes B is the dominant shape with one A. In addition, note that there are only 4 possible YES items AA, BB, CC and DD. In order to insure an equal number of YES and NO items are presented, YES items were presented more frequently than NO items.

Including all 25 shape families, there are 100 (4 x 25) YES items and 300 (12 x 25) NO items. Each of the 400 shape combinations was assigned an Item Number.

Each session of 250 items included an equal number of both types of items. Thus, YES items are seen 125 times per session and NO items are seen 125 times per session. Because there are fewer YES items to distribute, YES items were seen approximately 3 times more often than NO items.

Following a short reaction time test, subjects completed 10 sessions of 250 items per session for a total of 2500 items per subject. Subjects were encouraged to take breaks between sessions and were paid the full amount only if they completed all 10 sessions. Five different fixed item orders were used so each subject saw the same order two times (Figures 6 and 7). Over 250,000 lines of response data were generated in Aim 1. Note that, while there are 400 possible shape combinations as reflected in the orders, there are far more possible item presentations because the location and orientation of each shape on the checkerboard is selected randomly. Investigators postulated in advance that the random differences caused by item location and orientation was of minimal impact compared to the actual shape selection.

Subject ID	Order Number	Line Number	Item Number	Response Time	Subject Response	Is Response Correct?
1001	1	1	69	2.4166	0	1
1001	1	2	323	3.3333	1	1
1001	1	3	393	3.6666	1	1
1001	1	4	354	2.1666	1	1
1001	1	5	82	1.2500	0	1
1001	1	6	95	1.4166	0	1
1001	1	7	375	3.6666	1	1
1001	1	8	106	3.5833	1	0
1001	1	9	306	5.5000	1	1
1001	1	10	20	2.2500	0	1
1001	1	11	372	4.3333	1	1
1001	1	12	31	2.7500	0	1
1001	1	13	376	5.7500	1	1
1001	1	14	319	5.4166	1	1
1001	1	15	386	2.7500	1	1

Fig. 4, Shown above is an extract of the first 15 Aim 1 responses from subject 01. The Item Number column shows the designated number for the shape combination. Line number 8 shows an incorrect response for Item Number 106. The items were presented in 5 different fixed orders.

Lab facilities at the Stanford University Veterans Administration Hospital in Palo Alto, California, were used for the Aim 1 experiment. Subjects were drawn from the Bay Area's diverse population and a representative sample was recruited. Only one subject was not able to complete all ten sessions. However, results from 15 subjects were discarded because they showed a high error rate (more than 20% incorrect) or for other reasons - some subjects fell asleep, etc.

Photo 2, Aim 1 Experiment at Stanford VA Hospital in Palo Alto, CA.

Subjects were instructed in how to take the test at the beginning of the process and they were shown a graphic of all shapes. The PI explained to each of the subjects that they did not need to memorize the shapes but that they should become familiar with some of the variations between shapes. The software displayed a green checkmark when the response was correct and a red arrow when it was incorrect.

Aim 1 Process

As indicated, the purpose of Aim 1 was to determine the relative difficulty of each of the 400 item combinations and to incorporate this information in the scoring algorithm to insure optimal test-to-test stability. Response time and proportion correct were the two variables used to quantify the difficulty level of each combination. Using Item Response Theory "IRT" a number of different statistical models were explored in analyzing the data. IRT proved to be highly efficient in establishing relative difficulty levels among the various shape combinations.

(Graphs on following page)

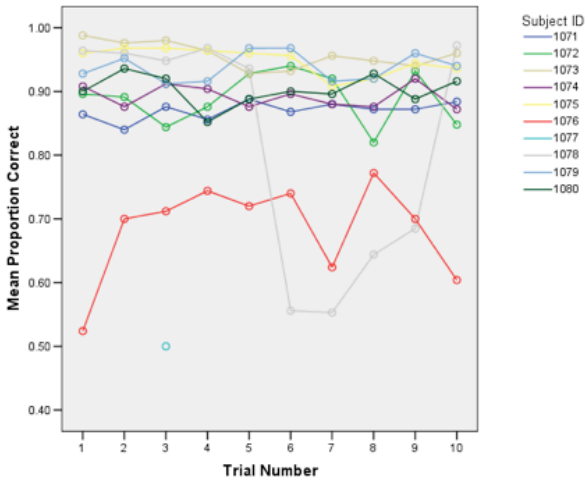

Fig. 4. Subject results showing significant drops in accuracy were eliminated. In this case Subjects 1076 and 1077 were eliminated based on low accuracy. Trials 6-10 for Subject 1078 were also eliminated because of data errors and low accuracy.

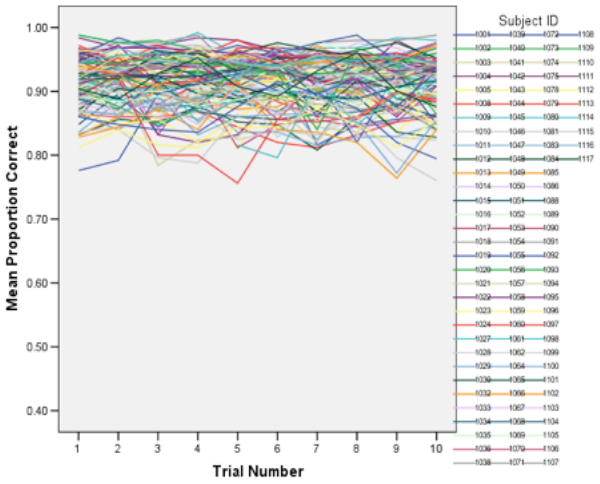

Fig. 5. Graph shows proportion correct (accuracy) by trial for each subject categorized as having “clean data”. Note that accuracy is above 80% for most subjects.

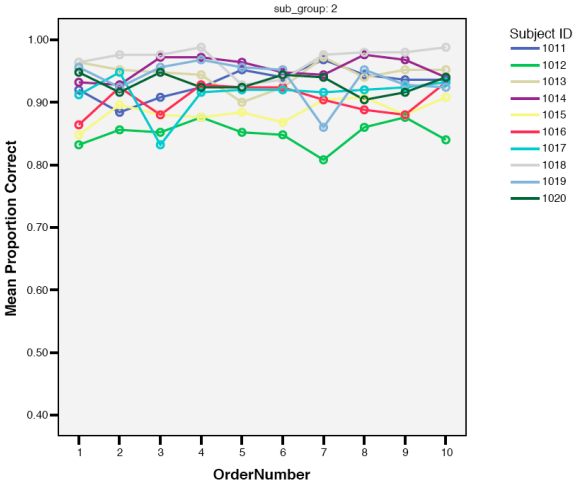

Fig. 6. Mean Proportion correct for different orders (see text).

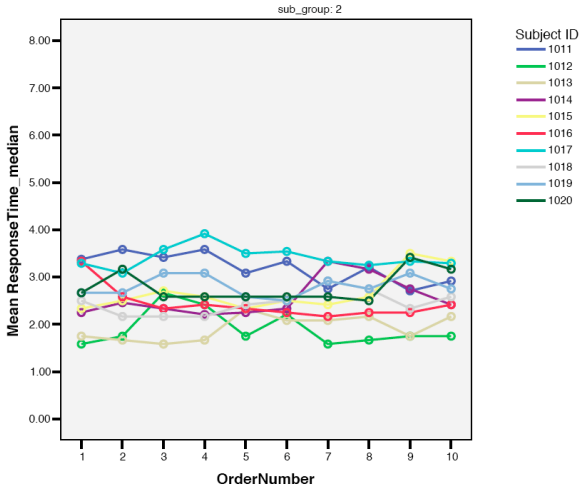

Fig. 7. Mean Proportion Correct and Mean Response Time and was constant for most subjects through all 10 sessions.

Experimenters observed a significant variation in speed and accuracy between subjects. However, individual subject results were mostly consistent from session to session. This would tend to confirm the hypothesis that each subject has an inherent or natural capability for maximum test performance. This concept supports the principle of using baselines to measure individual performance rather than external measures of absolute performance.

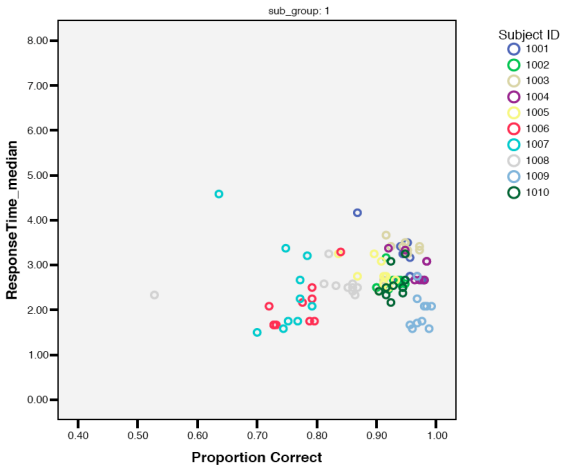

Fig. 8, Results from 10 subiects over 10 sessions.

Item Selection and Weighting

The investigators used graphs and IRT to explore possible solutions to the stability issue. It became apparent that items and item families had characteristics that could be selectively quantified. Mean Response Time and P Value (Proportion Correct) for all subjects are graphed below showing two example shape families (Figures 9 and 10). Each of the "different" items (item numbers beginning with "d") has 5 circles because each different item appeared in half of the 10 sessions. Each of the "same" items (item numbers beginning with "s") has 10 circles because they appeared on all 10 sessions.

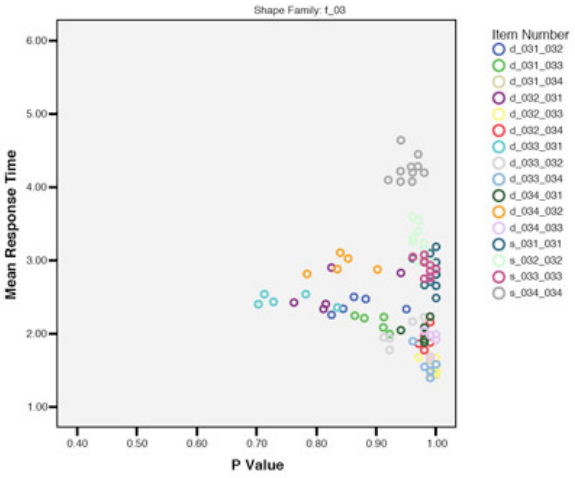

Fig. 9. Shape families showed similar positions on the P Value/Mean Response Time graphs.

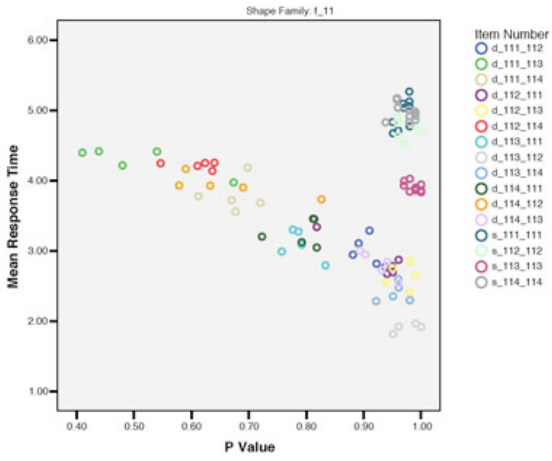

Fig. 10.

Item Eliminations?

Consideration was given to eliminating items that were very difficult for most subjects. A possible cut-off was 0.80 Mean Proportions Correct and to eliminate those items that subjects were able to answer correctly less that 80% of the time. This would eliminate the 57 most difficult items (out of 400). Nevertheless, it was decided to keep the existing item and shape library without these deletions for the remainder of project to preserve the integrity of the test for the entire experiment and because some investigators posited that difficult items might have psychological benefits and predictive value that was not fully understood at the time. (Fig.11)

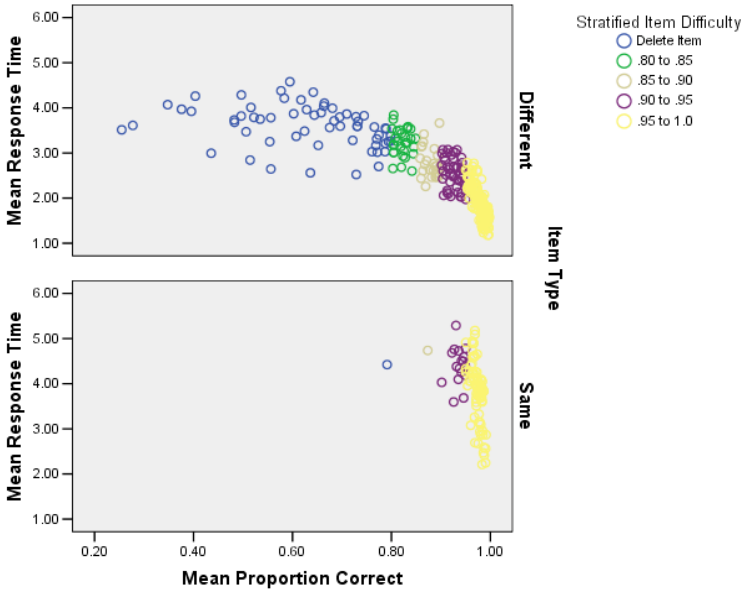

Fig. 11, Consideration was given to the deletion of some shape combinations. Ultimately, investigators decided to keep all shapes.

YES vs. NO Items

The Aim 1 data further quantified the difference in difficulty between NO items (one shape is different) and YES items (all shapes are the same). YES items showed a consistently higher mean proportion correct with a longer response time. NO items showed faster times but lower mean proportion correct. This would be consistent with subjects searching the screen of a NO item and quickly finding the different shape. With a YES item the subject might search the screen several times to be sure a different shape is not present, hence YES items tend to take longer to identify. Thus subjects who do not identify a different shape on a NO item in the first few seconds may respond quickly and incorrectly or spend additional time searching to be sure they are not guessing. Since subjects are given 7 seconds (Dwell Time) to respond, a conservative strategy would be to take the full 7 seconds searching at least until a different item is identified, otherwise assume it is a YES item and respond accordingly. Shorter or random dwell times might

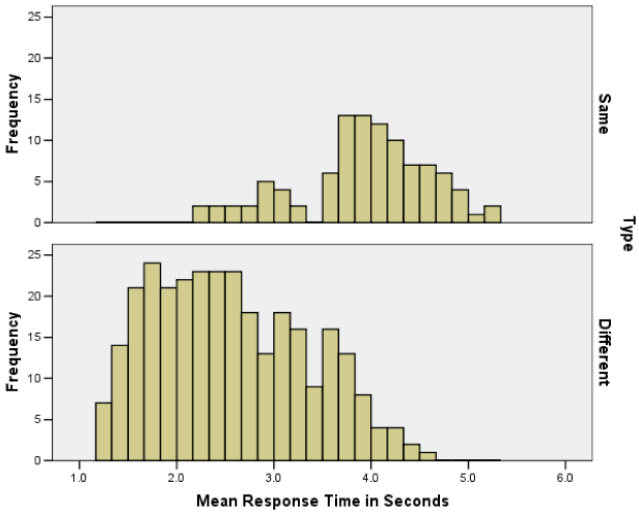

Fig. 12, Longer Mean Response times for YES (Same) items shows subjects searching longer on these items to be certain a different item was not overlooked.

undermine this strategy. Investigators determined to keep the 7-second dwell time and use IRT to include response time in determining item difficulty.

Aim 1 Results

Based on the Aim 1 experimental results the experimenters divided the shapes into groups or "bins" with items of similar difficulty in each bin. The test algorithm was then adjusted to draw a fixed number of items from each bin rather than randomly from the entire item pool. This insures that each test will contain an equal number of difficult and easy items.

Definition of the Bins

Definition of bins for NO (Different) Items was based only on IRT Difficulty. The cutoffs shown are assigned -2, -1, 0, 1, 2, and 3. The items are classified into 7 levels of difficulty ranging from very easy to very hard. Although these items were classified solely on IRT Difficulty, the item bins also increase in average Mean Response Time (RT). As shown below (Figure 13) IRT Difficulty and Mean RT are highly correlated for the NO items (correlation = 0.888).

By contrast, response time and IRT Difficulty were not highly correlated for the YES (Same) items (correlation = 0.477). Thus definition of bins for YES (Same) Items was based both on IRT and IRT Difficulty (Figure 14).

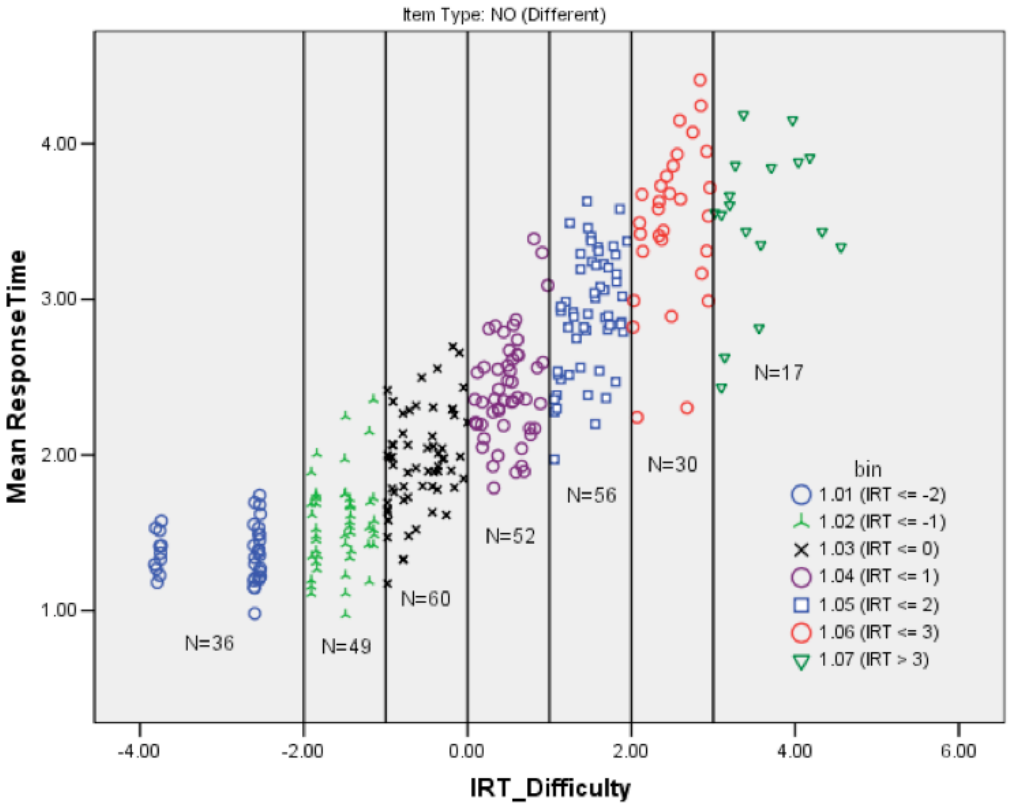

Fig.13. Scatter plot of Mean Response Time and IRT Difficulty for the NO (Different) Items. Vertical lines indicate bin assignments.

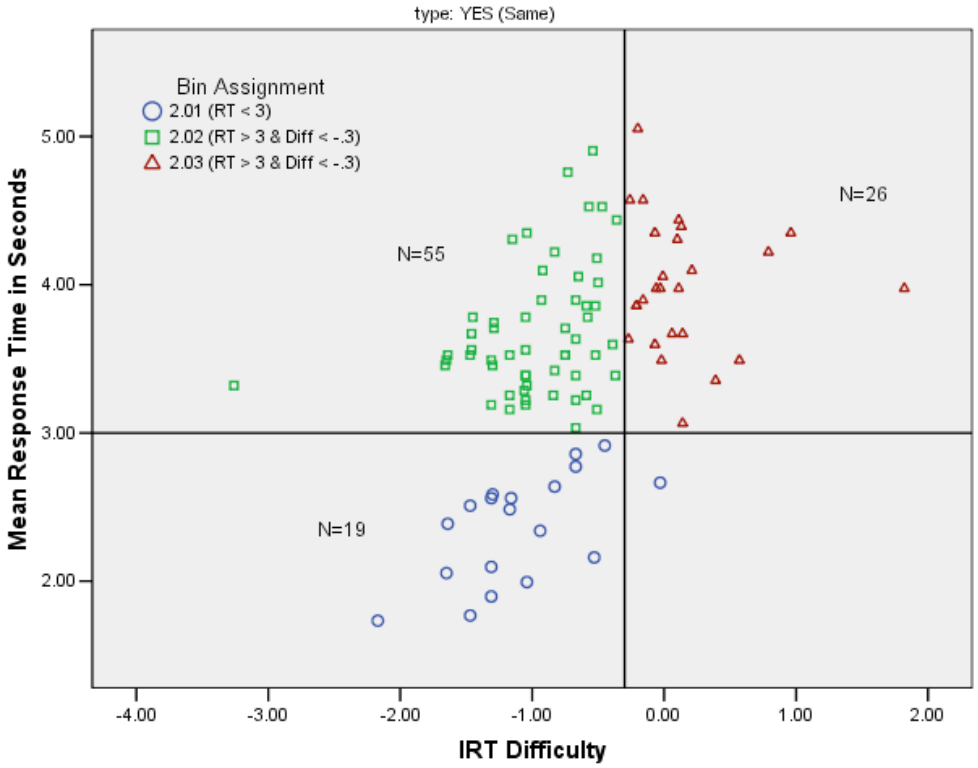

Fig. 14. Mean Response Time and IRT Difficulty for YES (Same) Items showing bin quadrants.

Bin assignments for YES Items is shown in table below (Note that single item in the forth bin was included in Bin 2.01 resulting in three (3) bins.)

Summary statistics for the YES Items.		
Bin 2.01:	Fast & Easy	Mean Response Time < 3.2 Seconds
Bin 2.02:	Slow & Easy	Mean Response Time > 3.2 Seconds and IRT Difficulty < -0.3
Bin 2.03:	Slow & Difficult	Mean Response Time > 3.2 Seconds and IRT Difficulty > -0.3

Once the bin definitions were established, each of the 400 possible shape combinations was assigned to a bin. The scoring algorithm for Aim 2 used this system to assign the same number of difficult and easy items to each test resulting in better test-to-test stability. It should be noted that from the users standpoint the test items appear to be randomly selected because the location and orientation of the shapes, the sequence of items and the selection of items within a given bin remains random.

Test Matrix

A test matrix was developed to determine how many items to select from each bin for a given test length. By using this matrix investigators are able to lengthen or shorten the test (with more items or fewer items) without significantly changing the difficulty level of the test. Items are drawn from the bins based on the percent of items in each bin.

Number and Percentage of Items per Bin for 50 and 80 Item Tests. The "standard" BLT test is 50 Items, 25 YES and 25 NO Items.

Item Type	Bin	Frequency	Percentage	Number on Test	Number on test
NO (Different)	1.01	36	12%	3	5
	1.02	49	16%	4	6
	1.03	60	20%	5	8
	1.04	52	17%	4	7
	1.05	56	19%	5	8
	1.06	30	10%	3	4
	1.07	17	6%	1	3
		300		25	40
YES (Same)	2.01	19	19%	5	8
	2.02	55	55%	13	22
	2.03	26	26%	7	10
		100		25	40

Aim 2 Completion Report

Aim 2 Background

The Aim 2 experiment followed a sleep deprivation protocol designed to determine the effectiveness of the BLT Alertness Test for measuring human fatigue. The trial was conducted at the Awake Institute, LLC. Lab in Arlington, MA. Fifteen paid subjects (aged 25-50 years) participated in the trial.

Aim 2 Procedures

The BLT test used in Aim 2 was the standard BLT Alertness Test with the revised software using the bin system as developed in Aim 1 for item selection. All response subject data, not just a computed score, was collected so that other scoring algorithms could be explored using different combinations of data sets and methodologies. Testers used were matched laptop PCs connected to the BLT server at AlertnessCentral.com. To motivate subjects and maintain realism a simplified scoring system was used to compute a test score at the conclusion of the test. This was displayed with a result graphic (Fig. 15) showing current test performance and the four previous scores.

With the exception of two short server breakdowns, the server based system performed well and afforded the investigators the opportunity to track and download data in real time in different locations on the West and East Coasts.

The data from the trial were analyzed to determine the general effectiveness of the test as a measure of fatigue induced impairment. This involved using IRT and other statistical methods. Finally, different scoring algorithms were examined to explore correlations between physiological data, subjective data, projected circadian patterns and BLT data.

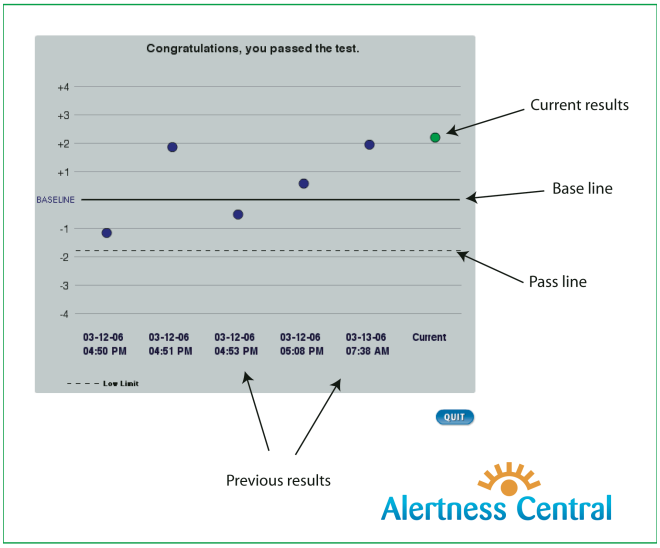

Fig. 15. Graphic displayed at the end of each test (shown here with explanatory arrows which are not part of the regular display).

Aim 2 Process

Subject Protocol

During the Aim 2 sleep restriction experiment, subjects were required to stay in the laboratory for two consecutive days. They reported to the laboratory at 9 a.m. on Day 1. The first three hours were used for setup (EEG wire-up, additional practice of BLT and Performance test) and for establishing their individual BLT baseline. Test sessions (see below) of one-hour duration were initiated every 2 hours from 1200 hours to 0600 hours on Day 1 and Day 2. No sleep was allowed during or in between test sessions on Days 1 and 2. Three hours of sleep were allowed in the morning of Day 2 (0900 hours to 1200) prior to starting the first experimental session at 1300 on Day 2. No caffeinated beverages were allowed at any time of the experiment.

On each test day, subjects completed ten bi-hourly test sessions (starting at 12:00). Each test included several subjective alertness/mood tests (e.g., Visual Analog Scales, Thayer Activation-Deactivation Adjective Checklist, Karolinka Sleepiness Scale), performance tests (5-min performance vigilance task PVT, 25-min driving simulation task, 50-screen four-choice reaction time test), and four BLT Alertness Tests.

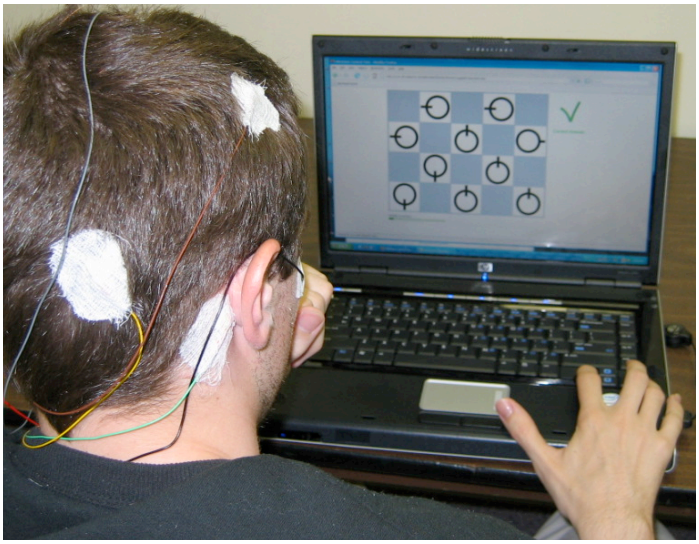

Photo 2. Subject during Aim 2 experiment.

Sequence of Subjective, Performance and BLT Tests

Each one-hour test session followed the following sequence:

	Test	Description
1	Subjective tests	Questionnaires (Thayer Mood Scale, Karolinska Sleepiness Scale, Visual Analog Scales)
2	BLT tests	2 Standard 50 item BLT
3	Simulator driving	25min computer based driving task Supplier: Systems Technologies, Inc.
4	Subjective tests	Questionnaire (Visual Analog Scales)
5	2 BLT tests	2 Standard 50 item BLT
6	Four-Choice Reaction Task	Test adapted from Wilkinson and Houghton 1975, Test presents a spot at one of 4 locations on the computer screen.
7	Subjective tests	Questionnaire (Visual Analog Scales)
8	Psychomotor Vigilance Task (PVT)	5-min simple reaction time task Supplier: Ambulatory Monitoring, Inc.
9	Subjective tests	Questionnaire (Visual Analog Scales)

Note: The PMI FIT 2000-3 fitness-for-duty/impairment screener that measures eye reflex movement as a fatigue indicator was also used with some subjects on an experimental basis.

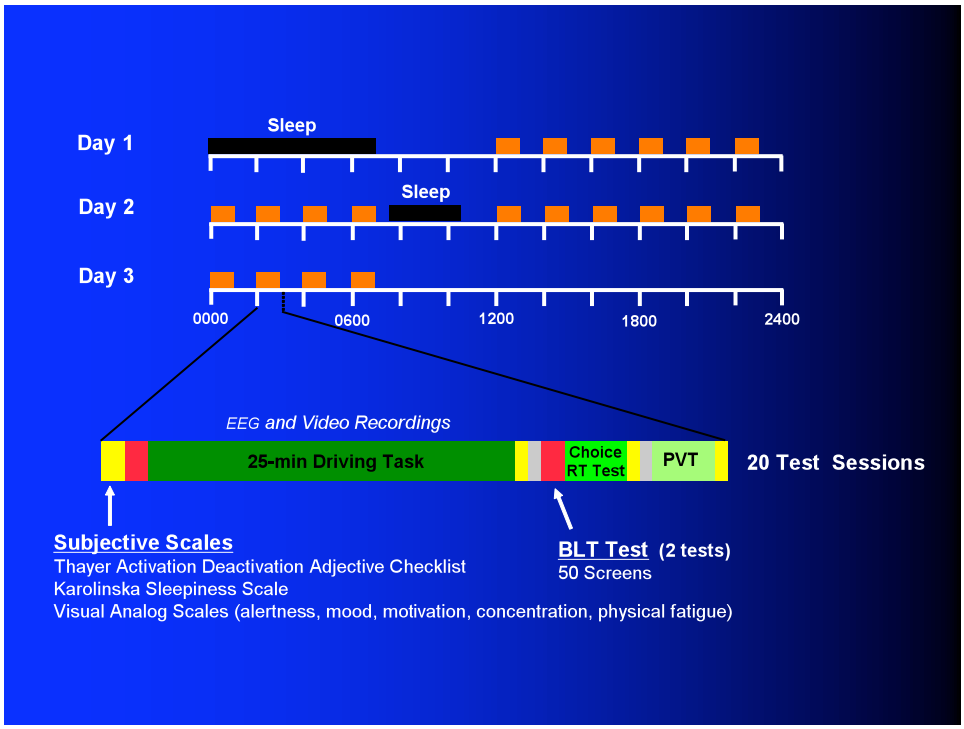

Fig. 16. Graphic showing sequence of testing during the Aim 2 experiment.

BLT Baseline

Each subject took the BLT Test at least 20 times for advanced practice on a separate day during the week prior to the experiment. Baselines were then established by taking the test 10 times. Individual differences are indicated in baseline performance in Table 000.

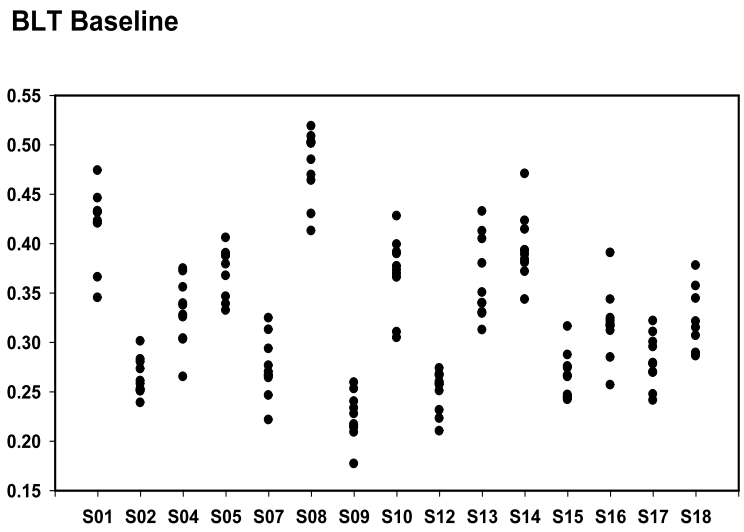

Fig. 17. Subject baseline performance for the Aim 2 experiment.

"Gaming"

Most subjects show results with a high proportion of items correct (> 80%) so a sequence of incorrect responses usually is an indication of inattention or deliberate "gaming" of the system. The software provides a setting that requires subjects to restart the test after scoring three incorrect responses in a row. For the experiment subjects were allowed to continue with the test after a warning box appears on screen rather than having to restart. One possible gaming strategy, therefore, is to respond randomly or very quickly and override the warning message. Subjects who did get many three wrongs in a row were counseled to see if they understood the test or if they were having other problems. In actual use the test setting requiring a restart eliminates this particular gaming strategy.

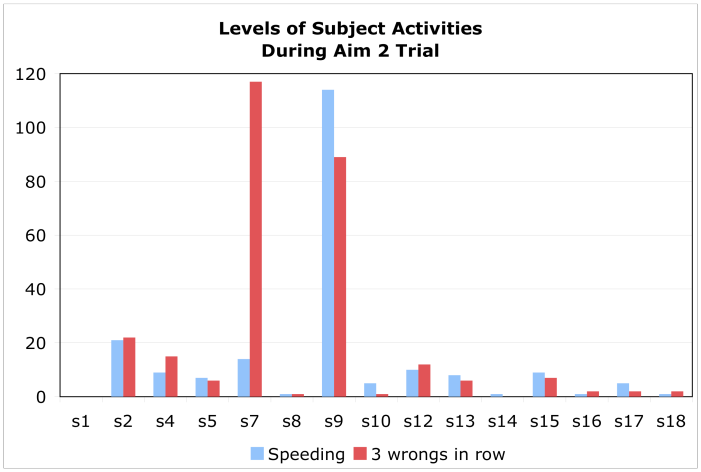

Fig. 18. Two subjects showed evidence of "gaming" the test. In actual use this possibility would be eliminated by a change in test settings.

Subjective Data

Subjective data was obtained from subject questionnaires administered after each set of tests using a variety of Visual Analog Scales (VAS). This VAS block used of five scales to measure: arousal, mood, motivation, concentration and physical fatigue.

Measured	VAS Scale
Arousal level:	0=very sleepy - 100=very alert
Mood:	0=very bad mood - 100=very good mood
Motivation:	0=not motivated at all - 100= very motivated
Concentration:	0=unable to concentrate - 100=able to concentrate very well
Physical fatigue:	0=very fatigued - 100=not fatigued at all

During each bi-hourly test session, subjects completed the VAS block four times. The first and second VAS block of each test session was immediately followed by a block of two consecutive BLT tests (BLT pair). These blocks with the subsequent BLT pairs were separated by the driving task. The third and fourth VAS blocks were conducted later during the test session, before and after the PVT test.

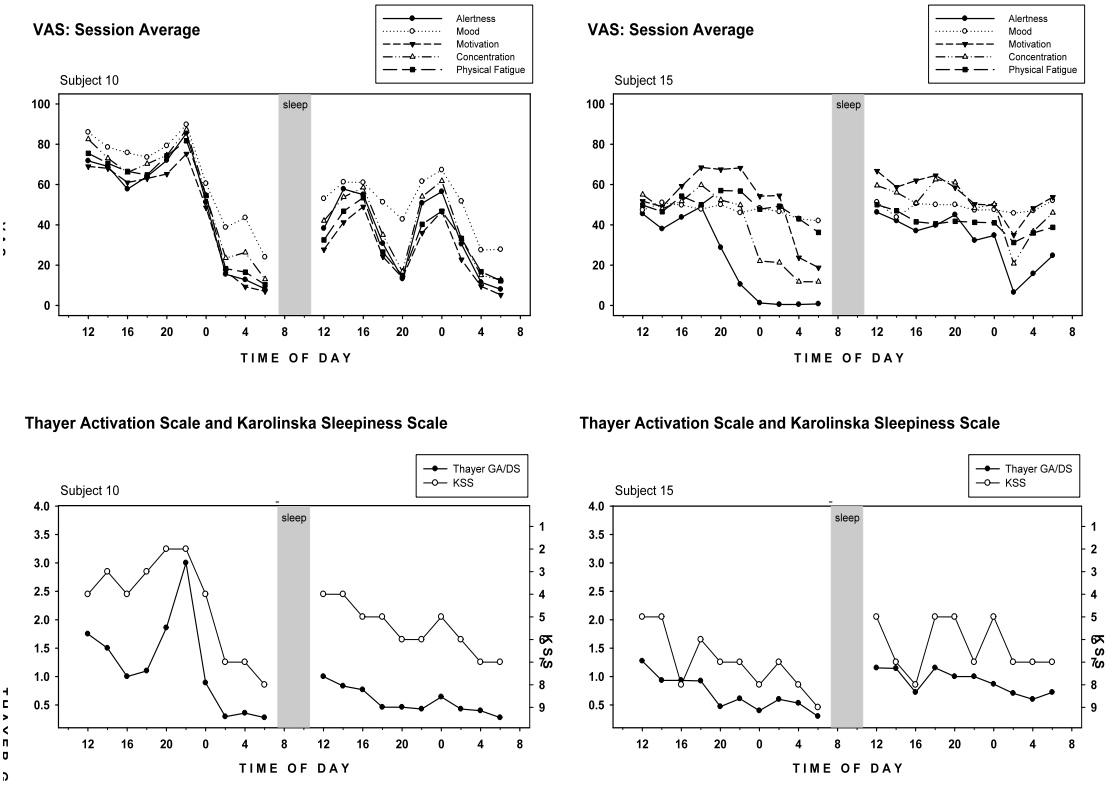

Fig. 19. Charts showing indicated subjective feelings of sleepiness for selected subjects during the Aim 2 experiment.

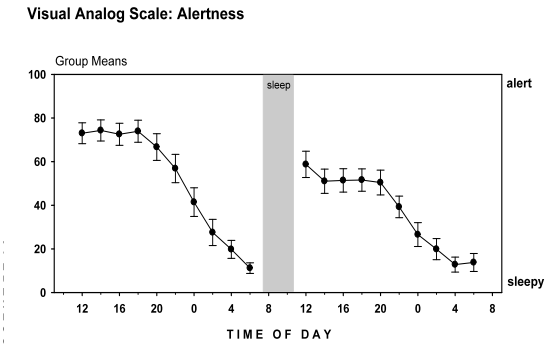

Fig. 20. VAS Graph- Group Mean, Alertness.

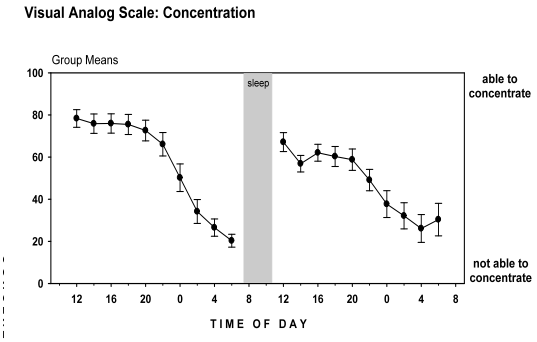

Fig. 21. VAS Graph – Group Mean, Concentration

VAS subjective data correlated well with circadian performance expectations as indicated by the Group Means charts (Figures 20, 21). This would indicate that the sleep deprivation protocol was succeeding in making the subjects feel tired. Note that there was some variation on an individual level. This was an expected indicator that different subjects react to sleep deprivation differently.

PVT Data

The PVT device is a well-studied measure of reaction time and vigilance. Test results were consistent with expected circadian patterns with some individual variation. (Figure 22).

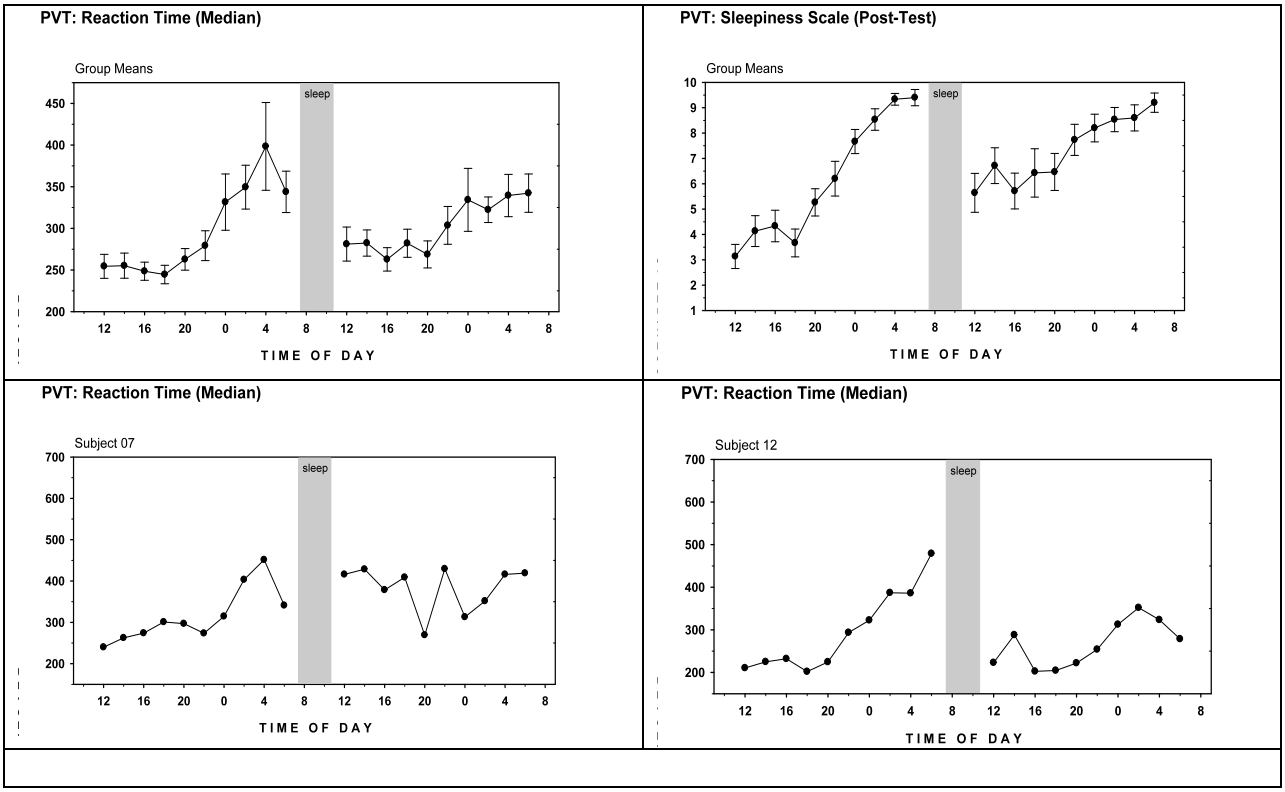

Fig. 22. PVT graphs Group Means (upper), sample individual results for Subjects 07 and 12.

BLT Test Data

Using the basic BLT scoring formula (before alternative scoring formulas were tried) the BLT Test showed a general correlation with the VAS and PVT and anticipated circadian patterns.

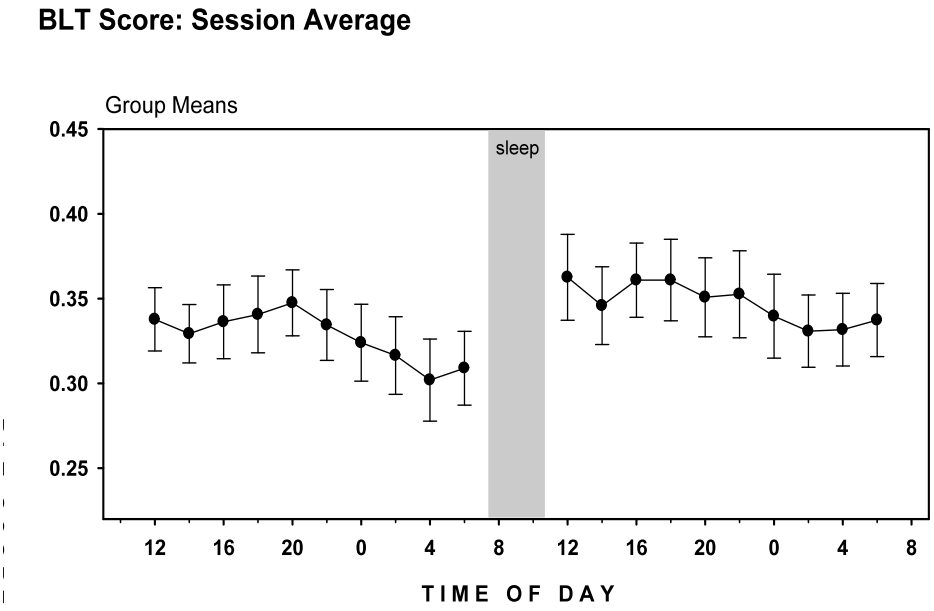

Fig. 23, BLT Test results, Session Average, Group Means.

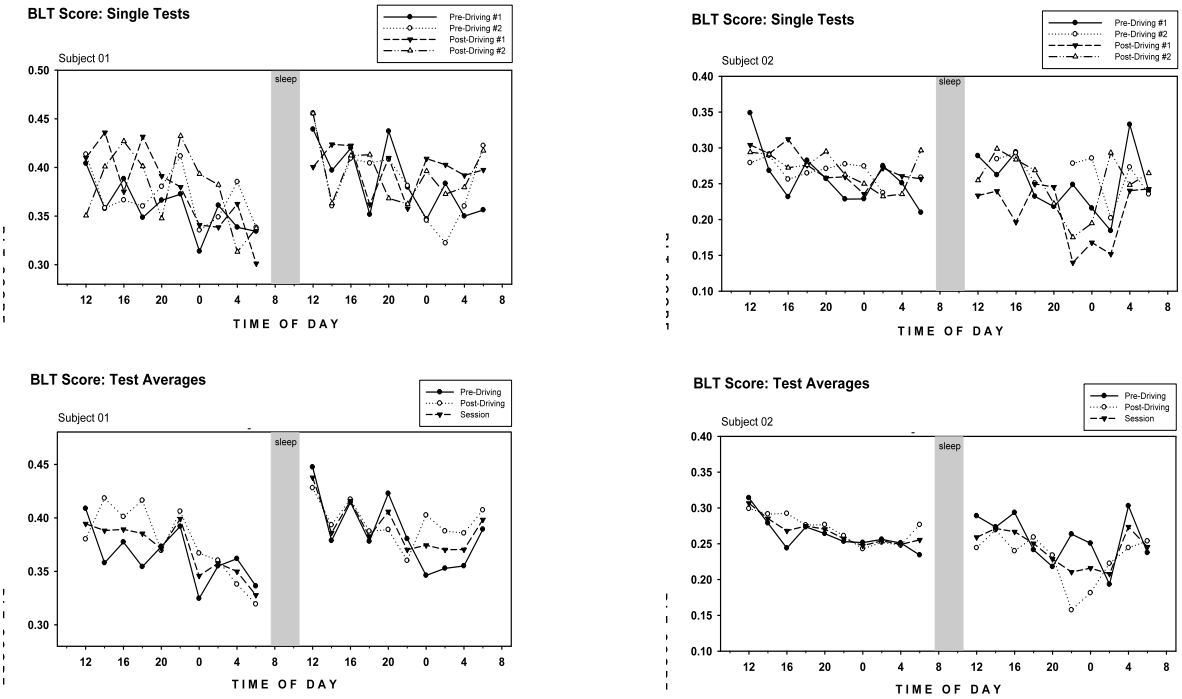

Fig. 24. BLT Test scores for Subjects 01 and 02 showing single test score results (upper) and test averages (lower).

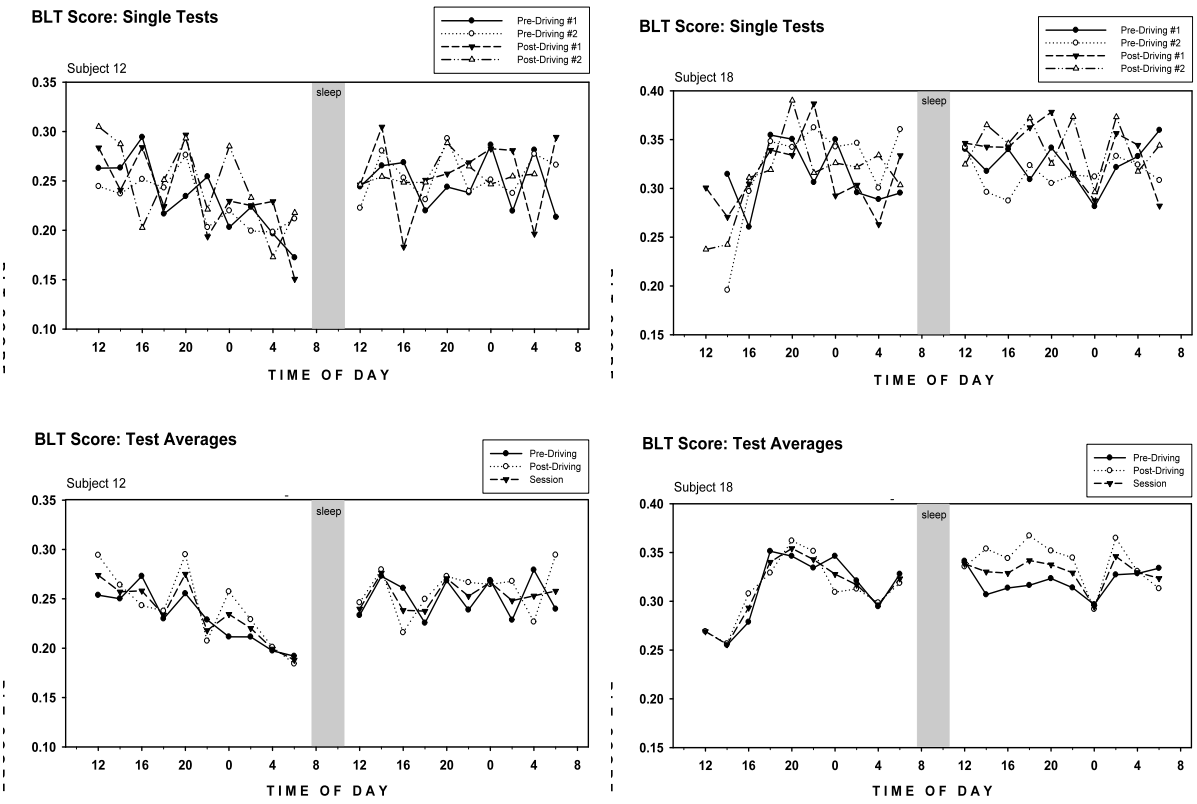

Fig. 25. BLT Test results for Subjects 12 and 18.

BLT Score Components

Subject components were analyzed to explore possible avenues for scoring development.

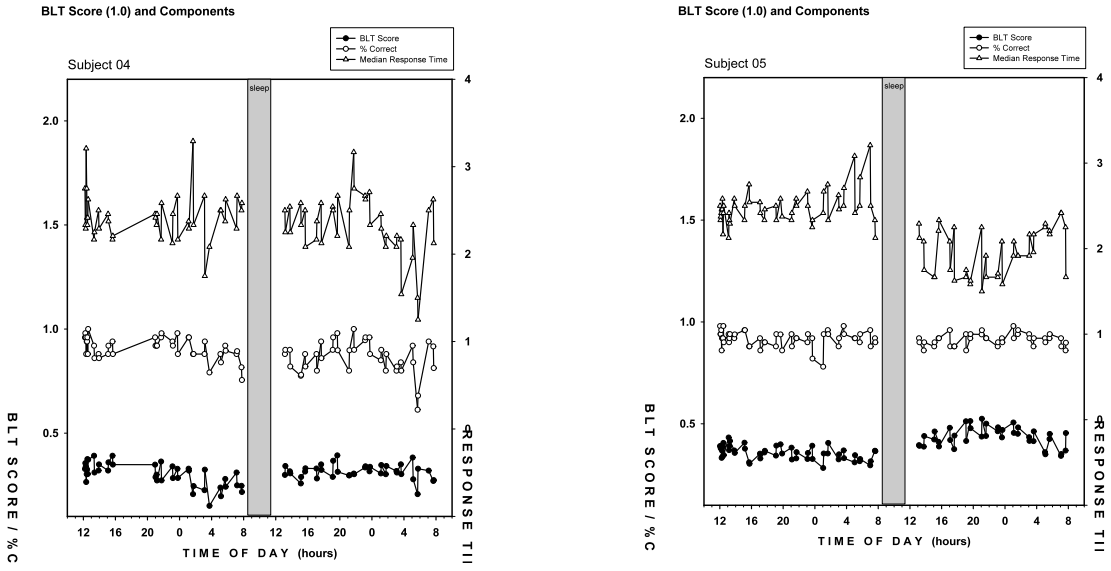

Fig. 26, BLT scores component graphic analysis. Graphic shows different score components for 2 subjects (S04, S05) during the Aim 2 experiment. Top line shows the BLT Standard Score, middle line is % Correct, lower line is Median Response Time.

Optimizing the Scoring Algorithm

Using statistical and graphic tools the investigators experimented with a number of scoring algorithms to maximize predictive value of the test. The goal was to find a scoring algorithm that showed the best correlations between BLT data and other measures including subjective measures.

With data from Aim 1 it was possible to add scaled item difficulty as one of the potential variables along with response time and accuracy. In addition, the investigators considered algorithms that measured correct and incorrect responses differently or that treated YES and NO responses differently. For example, investigators examined whether counting just YES items or NO items by themselves might show better correlations. The optimal penalty for incorrect responses was also considered for its effect on calculated score. Relating the relative importance of speed and accuracy was apparent from the start of the experiment and the developed formulas each supported a different approach to how best to weight these factors. These approaches were narrowed to 6 different scoring algorithms:

Score	Numerator	Denominator
Team01	Sum of Scaled Difficulty for Correct Responses – W * Number Wrong Where W=2	Median Response Time for Correct Responses * Number Correct
Team02	Sum of Scaled Difficulty for Correct Responses – W_S * Number Same Items Wrong - W_D * Number Different Items Wrong Where W_S=2.1 and W_D =1.9	Median Response Time for Correct Responses * Number Correct (same as Team01)
Team03	Same as Team01, except that the difficulty values are scaled from .5 to 1.5 (instead of from .9 to 1.1). W=2	Median Response Time for Correct Responses * Number Correct (Same as Team01)
Team04	Same as Team 02, except that the difficulty values are scaled from .5 to 1.5 (instead of from .9 to 1.1) and **W_S=2.25 and W_D =1.75**	Median Response Time for Correct Responses * Number Correct (Same as Team01 and Team02)
Team05	Proportion Correct	Median Response Time for Correct Responses to Different/NO items * Number Correct NO items

Score	Numerator	Denominator
Team06	Same as Team04	Median Response Time for Correct Responses to Different/NO items * Number Correct NO items (same as Team05)

Table 1, below, shows correlations between individual subjective alertness scores (VAS) and for different BLT scoring algorithms. The results for the algorithm with the strongest correlation are highlighted for each individual. It appears that the optimal scoring algorithm may differ from individual to individual, which may reflect individual strategies.

Regression Model Results

Day 1 and Day 2 combined

	Standard Score (W = .5)	Team04	Team05	Team06	Proportion Correct	Standard Score W=2
s01	0.71	0.58	0.72	0.57	0.06	0.58
s02	0.27	0.35	0.18	0.46	0.74	0.61
s04	0.70	0.62	0.25	0.58	0.64	0.67
s05	0.33	0.21	0.27	0.14	0.43	0.23
s07	0.32	0.79	0.12	0.80	0.72	0.78
s08	0.19	0.24	0.59	0.40	0.05	0.15
s09	0.44	0.76	0.62	0.74	0.71	0.72
s10	0.59	0.43	0.65	0.60	0.19	0.59
s12	0.56	0.21	0.19	0.17	0.39	0.41
s13	0.74	0.40	0.39	0.40	0.23	0.62
s14	0.28	0.15	0.53	0.54	0.53	0.34
s15	0.55	0.53	0.41	0.43	0.43	0.51
s16	0.76	0.83	0.32	0.55	0.79	0.87
s17	0.19	0.30	0.39	0.35	0.01	0.02
s18	0.01	0.12	0.61	0.46	0.00	0.10

Validity of the BLT test was assessed using correlations between various testbed parameters and BLT measures. The strength of the correlations varied between individuals. Table 1 below shows correlation coefficients for one selected subject, using two selected BLT scoring algorithms. Using session means for the various parameters, the individual correlations between BLT scores and testbed parameters were significant for the subjective (VAS, Thayer) and objective (PVT, driving) testbed parameters.

Reliability of the BLT test was assessed using correlations between two consecutive BLT tests. Table 2, below, shows correlation coefficients for each subject (excluding two subjects with speeding problems), using different BLT scoring algorithms. Reliability was relatively good for several subjects with significant correlations, but for other subjects, test reliability was not as strong.

Testbed Parameter	Score 5 (Session Mean)	Score 5_5 (Session Mean)
Subjective Measures		
VAS Alertness (Session Mean)	0.820 ***	0.665 **
VAS Mood (Session Mean)	0.913 ***	0.836 ***
VAS Motivation (Session Mean)	0.865 ***	0.743 ***
VAS Concentration (Session Mean)	0.797 ***	0.645 **
VAS Phys. Fatigue (Session Mean)	0.716 ***	0.591 **
Thayer GA/DS	0.644 **	0.590 **
PVT Measures		
Mean RT	-0.809 ***	-0.618 **
Mean SRRT	0.856 ***	0.713 ***
Lapses	-0.699 ***	-0.485 *
Driving Simulator Parameters		
Off-Road Accidents	-0.612 **	-0.503 *
Centerline Crossings	-0.823 ***	-0.731 ***
Missed Divided Attention Task	-0.750 ***	-0.579 **
Divided Attention Task RT	-0.877 ***	-0.689 ***
Lane Deviation Variability	-0.802 ***	-0.636 **
Steering Wheel Variability	-0.746 ***	-0.562 **
Heading Error Variability	-0.792 ***	-0.632 **
Curvature Error Variability	-0.750 ***	-0.570 **

Table 1

Significance Levels: *** p<0.001, ** p<0.01, * p<0.05

Subject	Score 5	Score 5_5	Score 1_5
S01 (n=40)	0.415 **	0.569 ***	0.548 ***
S02 (n=37)	0.303	0.307	0.277
S04 (n=34)	0.212	-0.101	0.035
S05 (n=40)	0.751 ***	0.822 ***	0.780 ***
S08 (n=39)	0.233	0.453 **	0.367 *
S10 (n=39)	0.535 ***	0.641 ***	0.597 ***
S12 (n=39)	0.433 **	0.403 *	0.359 *
S13 (n=38)	0.638 ***	0.457 **	0.491 **
S14 (n=39)	0.268	0.550 ***	0.448 **
S15 (n=40)	0.752 ***	0.743 ***	0.740 ***
S16 (n=36)	0.282	0.318	0.238
S17 (n=40)	0.219	0.284	0.228
S18 (n=39)	0.298	0.418 **	0.310

Table 2

Significance Levels: *** p<0.001, ** p<0.01, * p<0.05

Aim 2 Results

Results Summary

The various test bed measures (subjective tests, PVT, four-choice test and driving task) showed the expected circadian trend with consistently impaired levels at night. SRT scores were significantly lower at night (0200, 0400, 0600) as compared to the two test sessions with the highest scores of each respective test day. On the group level, the SRT score (standard scoring algorithm 1.0) correlated well (Spearman R>0.8) with many of the test bed measures. However, on the individual level, correlations were less strong and varied greatly between subjects.

The significant result of the Aim 2 study was the derivation of a new and unexpected scoring algorithm. This algorithm (a derivation of Team 5) places greater emphasis on proportion correct. Future versions of the BLT test will use this derived formula.

Aim 3 Completion Report

Aim 3 Background

Workplace testing was essential to insure the usability goals had been met and to better understand operators' reaction to testing. The Emergency Department at the University of Maryland Medical Center (Photo 3) was selected because the interns in the department work have shiftwork and long hours with a high risk for fatigue.

Photo 3. University of Maryland Medical Center.

Aim 3 Procedures

Twenty physicians participated in testing on ten consecutive work shifts. Test sessions were conducted before, during and after each shift and included subjective alertness/mood tests and the four-choice reaction time test (PVT). Participants also completed sleep diaries and daily operational performance questionnaires, and post-study operational feasibility questionnaires. Testing was done with laptop PCs placed in a private room. Software used was the standard BLT Alertness Test software running on the BLT Internet-based server. After the subject testing was done, supervisors completed an operational feasibility questionnaire.

Aim 3 Process

All testing was conducted on site with staff members acting as local managers and ten supervisors sequentially involved during the various shifts. Each was provided with a short Handbook explaining how the system worked, how to sign up participants and how to review test results. Subjects received a Subject Handbook with test instructions and a graphic of the entire library of shapes used on the test. There were no significant problems with understanding the test or in managing the process.

The PI filed a Weekly Summary Report and the on-site Administrator a weekly, or more frequent, report with email distribution to each investigator highlighting subject issues and general performance. Testing began on June 8 and ended on November 4, 2008.

"Activities"

In tracking test results particular attention was paid to "activities" to include speeding and guessing as indicators of inattention. Three subjects s12, s14 and s25 showed higher activity levels (over 24) but most subjects had less than 15 instances of speeding and guessing combined. Since subjects took the test on average 70 times during the trial these activity levels indicate that most subjects were cooperating. Subjects with high speeding levels were questioned after the test. Apparently, they had concluded that the scoring system favored speed over accuracy and were responding as fast as they could.

Aim 3 Results

Aim 3 was focused on issues of implementation, employee acceptance and general practicality. Data was collected with questionnaires for both subjects and supervisors.

Aim 3 Questionnaire Results – Supervisors

The Supervisors Operational Feasibility Questionnaires completed by 10 supervisors were generally positive toward the concept of alertness testing. Most indicated that alertness/impairment testing would help increase employee awareness of alertness/performance on the job. ("Yes"-7 supervisors, "somewhat" -3 supervisors). The majority of supervisors rated employee's acceptance of alertness/impairment screening as "favorable" (2 supervisors), "open to the concept" (3 supervisors), or "neutral" (3 supervisors). Two supervisors thought employees would reject the concept. The majority of supervisors rated the potential benefits of alertness/impairment screening as "beneficial" (7 supervisors) or "somewhat beneficial" (2 supervisors). One supervisor marked "not sure". Seven supervisors were "somewhat concerned" about privacy, three supervisors were "not worried".

Questionnaire Results – Subjects

1) 16 (of 20) subjects felt that the study increased, or somewhat increased, their awareness of alertness/performance on the job.
2) Except for one subject, participants did not feel that the testing had any effects on their personal sleep habits and/or other actions to improve their on-the-job alertness. A few subjects said they had a little less sleep because the study testing increased their time at work.
3) 8 (of 20) subjects thought the BLT test was sensitive, or somewhat sensitive, to impaired alertness/performance levels. Four subjects were not sure, and eight subjects thought it was not sensitive.
4) 14 (of 20) subjects rated the design of the BLT test as very or somewhat suitable for keeping the user interested in the task throughout the test and across repeated testing over consecutive work days. Two subjects were not sure, and four subjects thought it was not suitable.
5) For 11 (of 20) subjects, it was not a problem, or usually not a problem, to complete the BLT test during the shift. For six subjects it was a problem sometimes, and for three subjects most of the time.
6) 9 (of 20) subjects thought alertness/impairment screening would encourage, or somewhat encourage, good sleep habits and other actions by employees to improve on-the job alertness. 11 Subjects thought it would not help.
7) However, more subjects (16 of 20) thought alertness/impairment screening would encourage other actions by employers (e.g., better shift scheduling and/or other measures) to

improve on-the job alertness. Only four subjects thought it would not.
8) 11 (of 20) subjects thought they would feel better about their work environment knowing that all employees around them had been screened for alertness/impairment. Nine subjects thought it would not matter.
9) 14 (of 20) subjects thought privacy issues related to alertness/impairment screening would not worry them or would not matter. Six subjects were somewhat concerned.
10) 7 (of 20) subjects thought regular on-the-job alertness/impairment screening in their work environment would usually not be a problem. Eight subjects thought it would be a problem sometimes, and five subjects thought it would be a problem most of the time.
11) Only 3 (of 20) subjects rejected the concept of potential future alertness/impairment screening. The majority of subjects (17 of 20) said they would either favor the concept, be open or neutral to the concept.
12) Only 3 (of 20) subjects thought alertness/impairment screening had no potential benefits. The majority of subjects (14 of 20) rated it as very, somewhat or slightly beneficial. Two subjects were not sure, and one subject did not mark a response (verbal entry: "not much benefit").

Summary of Questionnaire Results

Questionnaire results showed a moderately positive response to alertness testing. Some concerns were raised about privacy, time taken and the accuracy of the results.

Alertness/Impairment Testing	**Subjects** (n=20)	**Supervisors** (n=10)
Potential Benefits	70% - Very/Somewhat/Slightly Beneficial	90% - Beneficial/Somewhat Benef.
Increase Employee's Awareness of Alertness/Performance on the Job	80% - Yes/Somewhat	100% - Yes/Somewhat
Acceptance by Employees	70% - Favor/Open-Minded/Neutral	80% - Favor/Open-Minded/Neutral
Concerns about Privacy	70% - Not Worried/Would Not Matter 30% - Somewhat Concerned	30% - Not Worried 70% - Somewhat Concerned

Data Analysis

The Aim 3 data set provided an opportunity to further explore validity and reliability of the BLT score, and further refine the scoring algorithm. Alertness tended to display the expected circadian pattern with lowest alertness at the end of the night shift and the beginning of the day shift (due to early start time). See Figures 27 and 28 for two selected subjects.

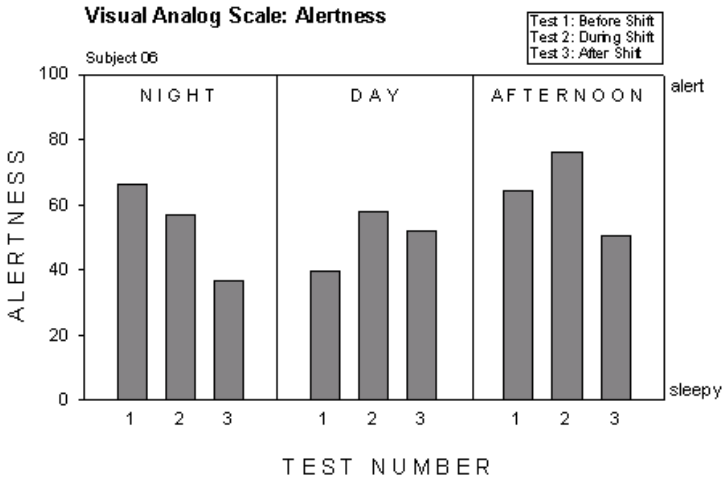

Fig 27, VAS results, Subject 06, Aim 3 Trial.

The basic BLT scoring formula applied to the BLT test data showed known problems of reliability with this formula (Fig. 29). That is, test to retest variability tended to mask the detected differences caused by subject circadian patterns. This does not mean, however, that in its basic form the BLT Test will not detect extreme impairment (beyond the usual circadian variation). Such extreme subject impairment was not encountered in the Aim 3 trial.

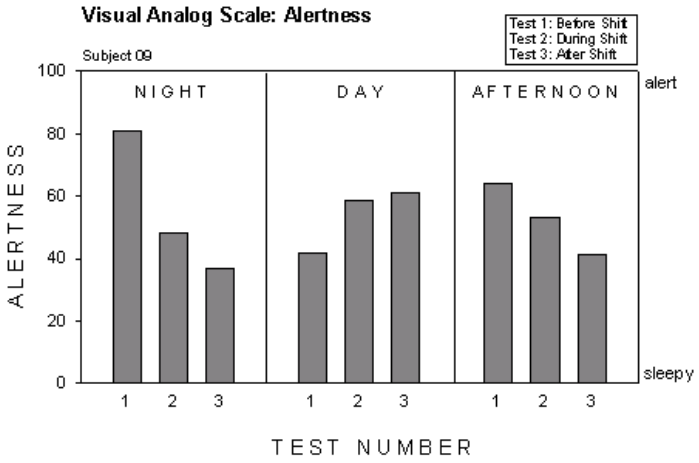

Fig 28, VAS results, Subject 06, Aim 3 Trial.

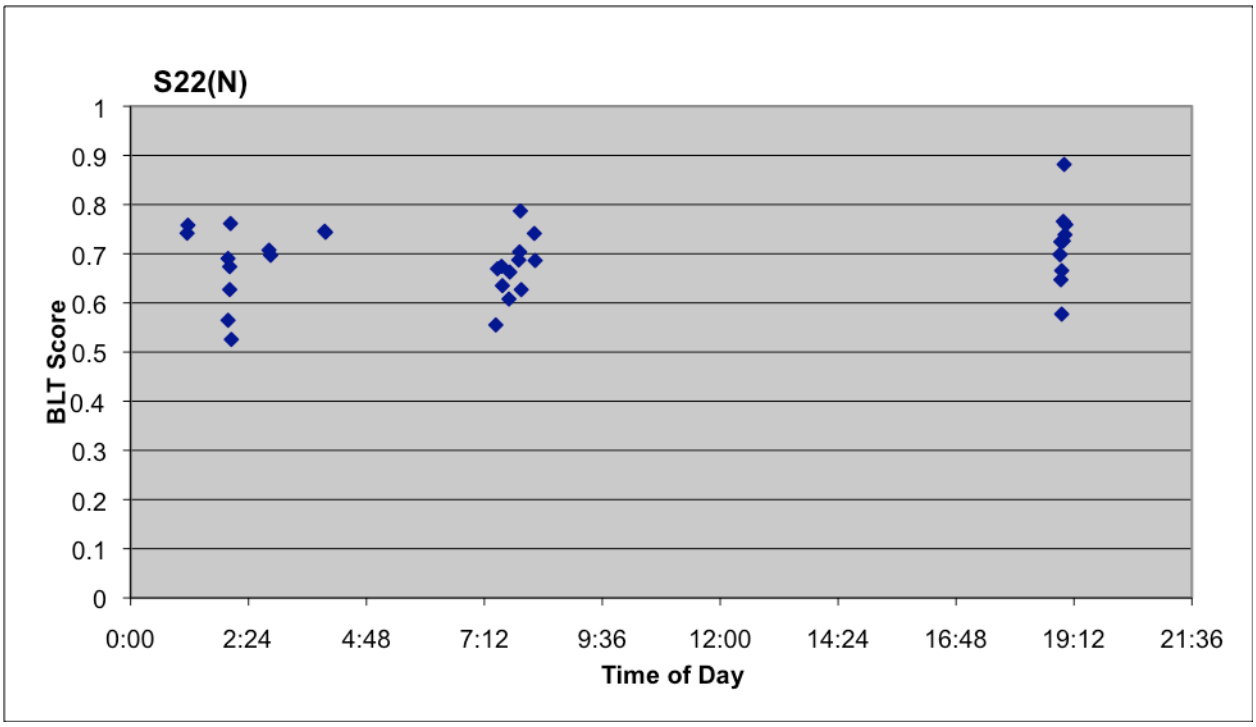

Fig. 29. BLT test score results for selected subject (S22) showing multi-day results.

In addition, correlations between expected circadian patterns and subjective feedback was compared to BLT test data using different scoring, see table below.

Component	Pearson Correlation Compare Night Begin And End	Pearson Correlation Compare Day And Night End	Sig Compare Night Begin And End	Sig Compare Day And Night End
Original BLT Score from Administration	-0.47	0.21	0.12	0.47
StandardScore	-0.48	0.11	0.12	0.71
Score01	-0.66	-0.69	0.02	0.01
Score02	0.56	0.64	0.06	0.01
Score03	-0.68	-0.72	0.02	0.00
Percent_Correct	-0.74	-0.77	0.01	0.00
Percent_Wrong	0.74	0.77	0.01	0.00
ResponseTime_sum	-0.58	-0.81	0.05	0.00
Mean Difficulty (Wrong Items)	-0.82	-0.76	0.00	0.00
Mean Difficulty (Correct Items)	0.24	-0.04	0.46	0.90
Median RT (Wrong Items)	-0.44	-0.77	0.15	0.00
Median RT (Correct Items)	-0.38	-0.70	0.22	0.01
Scaled_Difficulty_mean_Wrong	-0.82	-0.76	0.00	0.00
Scaled_Difficulty_mean_Correct	0.24	-0.04	0.46	0.90
Percent Wrong No(Different)	0.71	0.77	0.01	0.00
Percent Wrong Yes(Same)	0.65	0.09	0.02	0.77
Percent Correct No(Different)	-0.71	-0.78	0.01	0.00
Percent Correct Yes(Same)	-0.57	-0.07	0.05	0.82
Median RT Wrong No(Different)	-0.44	-0.74	0.15	0.00
Median RT Wrong Yes(Same)	0.69	-0.31	0.01	0.28
Median RT Correct No(Different)	0.02	-0.37	0.94	0.19
Median RT Correct Yes(Same)	-0.52	-0.68	0.08	0.01
ResponseTime_median_bin_1.01	0.32	-0.54	0.30	0.05
ResponseTime_median_bin_1.02	-0.40	-0.67	0.19	0.01
ResponseTime_median_bin_1.03	-0.57	-0.58	0.06	0.03
ResponseTime_median_bin_1.04	-0.56	-0.62	0.06	0.02
ResponseTime_median_bin_1.05	-0.60	-0.77	0.04	0.00
ResponseTime_median_bin_1.06	-0.13	-0.63	0.70	0.02
ResponseTime_median_bin_1.07	-0.16	-0.54	0.63	0.05
ResponseTime_median_bin_2.01	-0.03	-0.50	0.92	0.07
ResponseTime_median_bin_2.02	-0.53	-0.76	0.08	0.00
ResponseTime_median_bin_2.03	-0.46	-0.70	0.13	0.01
PercentCorrect_bin_1.01	-0.68	-0.63	0.02	0.02
PercentCorrect_bin_1.02	-0.64	-0.63	0.03	0.02
PercentCorrect_bin_1.03	-0.34	-0.68	0.27	0.01
PercentCorrect_bin_1.04	-0.68	-0.64	0.02	0.01
PercentCorrect_bin_1.05	-0.11	-0.43	0.73	0.13
PercentCorrect_bin_1.06	0.19	-0.23	0.55	0.43
PercentCorrect_bin_1.07	.	.	.	.
PercentCorrect_bin_2.01	-0.30	-0.06	0.34	0.84
PercentCorrect_bin_2.02	-0.45	0.29	0.14	0.32
PercentCorrect_bin_2.03	-0.42	-0.32	0.17	0.27

Table for single subject (s21) showing correlations with different scorings and potential score components. S21 shows relatively high correlations, some subjects showed no correlations.

While such correlation studies are not determinative of causal factors the process opens up potential future avenues for test analysis and future test design. The data from Aim 3 confirmed the success of new scoring options in improving reliability. The table below shows correlation coefficients for pairs of two consecutive BLT tests (one selected individual), using three different BLT scoring algorithms.

Correlation Coefficients (n=27)	**Reliability** (Consecutive Tests)
BLT Score 5	0.758 ***
BLT Score 5_5	0.758 ***
BLT Score 1_5	0.759 ***

Significance Levels: *** p<0.001, ** p<0.01, * p<0.05

Comments from the Subject Questionnaires demonstrated how quickly subjects could learn to adapt strategies that take advantage of any weakness in the algorithm. For example, in the Aim 3 trial many subjects learned that the basic BLT scoring algorithm tended to favor speed over accuracy. As a result some subjects chose to speed though the test rather than taking the time to carefully analyze each item. The better scoring algorithm that has been developed increases the weight given to accuracy. This algorithm is one of the major breakthroughs generated by the experiment and will benefit all future users of the BLT system.

Aim 3 Conclusions

Aim 3 achieved its aim of demonstrating that a properly designed computer-based alertness test is acceptable in a hospital ER workplace. Translating this finding to other workplaces will have to await further investigation and continued efforts toward commercial applications. Aim 3 also demonstrated the importance of good test design to prevent subjects from "gaming" the system. This was not an effort to evade the test, although in other environments this could of course occur, but more as a result of a natural tendency of human beings to use whatever systems are available to achieve the best possible results.

Test variability remains a significant issue that will be resolved through the use of more efficient scoring algorithms and/or better test designs. This problem was partially solved with the discovery of a better algorithm during these trials. The promise of a short graphic test for significant human impairment is an achievable goal - brought closer by this process.

Publications

Proceedings: Heitmann A, Bowles HM, Hansen K, Holzbrecker-Morys M, Langley TD, Schniptke, D [2009]: Seeking New Ways to Detect Human Impairment in the Workplace, International Conference on Fatigue Management in Transportation Operations, A Framework for progress, Boston, MA March 24 – 26, 2009.

Inclusion of Children Statement

No children were recruited or included in the subject pool.

Attachments:

Materials Available for other Investigators

We encourage other investigators to utilize the data collected during this experiment. Contact the PI, Dr. Theodore Langley Ph.D. at tlangley@bowles-langley.com or Henry Bowles at hbowles@bowles-langley.com.

Provided as separate files:

Final Financial Status Report

Inclusion Report

Final Invention Statement