In 2004, Bowles-Langley Technology, Inc. and Circadian Technology, Inc. conducted a sleep deprivation trial at the Circadian Sleep Lab in Lexington, MA. Funding for this research was from NIOSH. The results are detailed in the following report:

1 Report on Sleep Deprivation Trial

1.1.1 Sleep Deprivation Study (Aim 1)

The BLT Tester prototype was evaluated in a sleep deprivation protocol using Circadian Technologies' Alertness Testbed.

Ten healthy subjects aged 25-40 years were recruited after screening for medical problems and sleep habits (extreme morning or evening types were excluded). The gender and racial composition of these subjects are shown in Table 1.

	American Indian	Asian	Black	Hispanic	White	Other	Unknown	Total
Female	0	0	0	0	4	0	0	4
Male	0	1	2	0	3	0	0	6
	0	0	0	0	0	0	0	0
Total	0	1	2	0	7	0	0	10

***Table1**: Gender and Racial Composition of the Subjects in Sleep Deprivation Study*

Subjects participated in a 4-hour daytime training/practice session, and a 19-hour experimental test protocol - including one night of sleep deprivation (conducted a few days after the training/practice session). During the training/practice session, subjects had the opportunity to practice the BLT test and all other performance tests in order to minimize potential learning effects during the experimental protocol of the sleep deprivation study.

The overnight sleep deprivation study included a series of test sessions. Test sessions of one-hour duration were initiated every 2 hours from 1300 hours to 0700. The test protocol is illustrated in Figure 1. Each test session included multiple BLT tests, and the following Alertness Testbed measurements:

- Performance measures for three performance tests (driving simulation task, Psychomotor Vigilance Task (PVT), Wilkinson Four-Choice Test),
- Subjective measures (Visual Analog Scales and Thayer Activation-Deactivation Adjective Checklist),
- Frequency of behavioral signs of sleepiness (video recordings during driving task),

- Frequency of microsleep events (continuous recording of electro-physiological parameters).

Five BLT tests were conducted during each session, two tests before the driving task at the beginning of the test session, two tests after the 30-min driving task, and one test towards the end of the test session after the 10-min PVT. Multiple BLT tests were given in each test session in order to increase stability for validity testing, BLT session averages were used for statistical analysis (see below), and in order to provide data sets for reliability testing (see below).

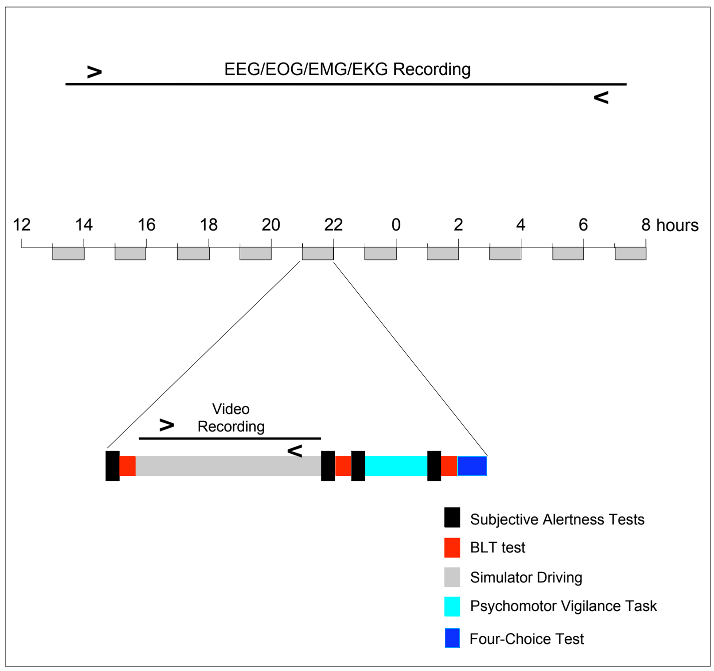

***Figure 1**: Experimental Test Protocol*

1.1.2 Statistical Analysis (Aim 2)

The data for the BLT and all Alertness Testbed measures were analyzed for each subject. As the subjects went through an overnight sleep deprivation protocol, the Testbed measures were expected to show a typical circadian pattern with lower alertness and performance during the night hours. Figures 2-10 show the Testbed results averaged across all ten subjects.

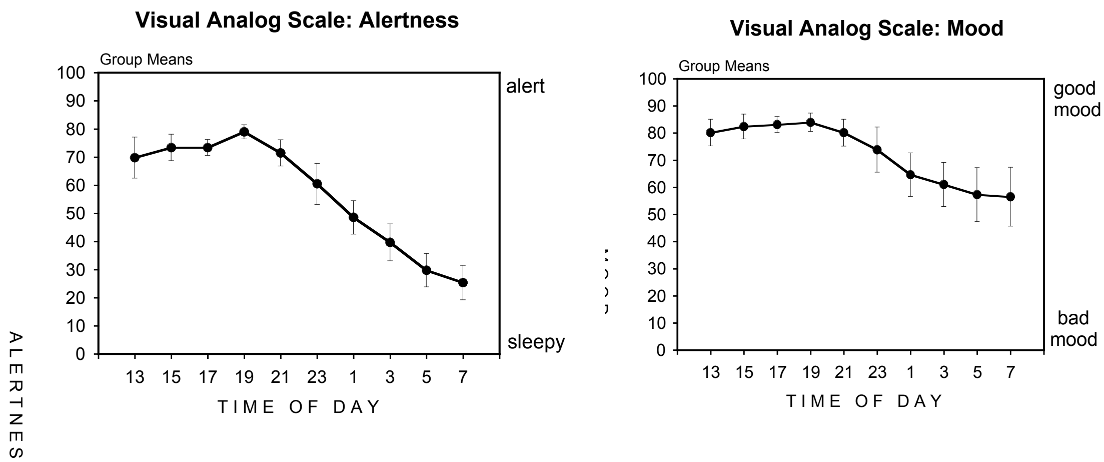

Figure 2: *Subjective Scales: Alertness (left) and Mood (right)*

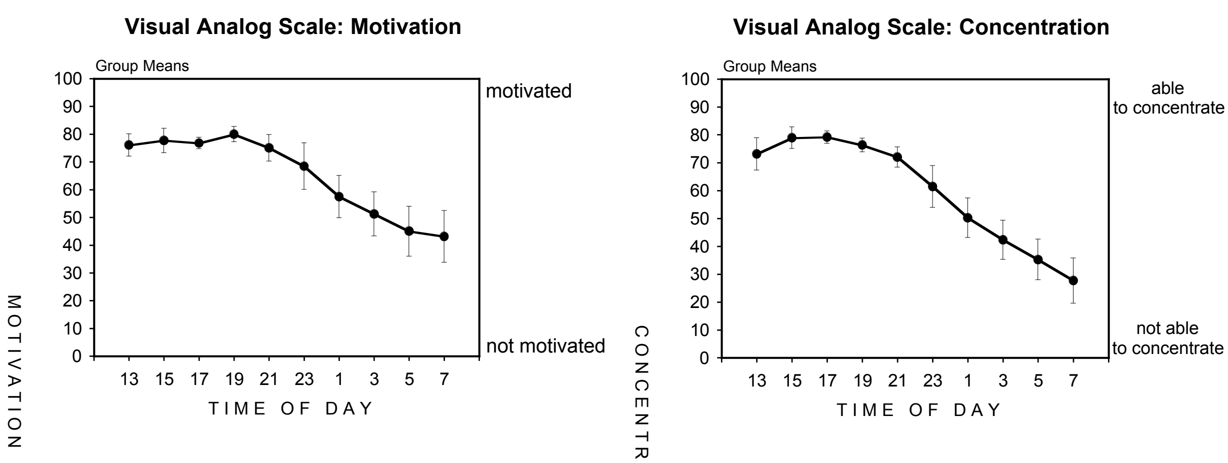

Figure 3: *Subjective Scales: Motivation (left) and Concentration (right)*

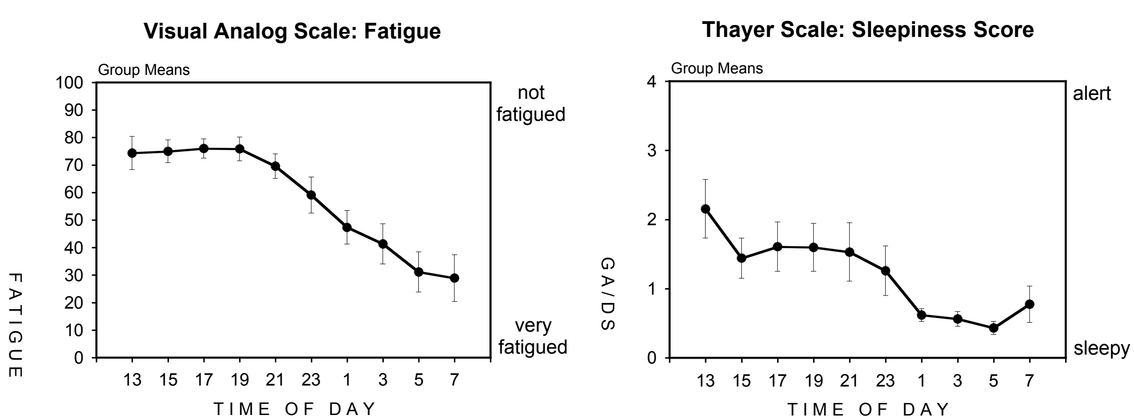

Figure 4: *Subjective Scales: Fatigue (left) and Thayer Sleepiness Score(right)*

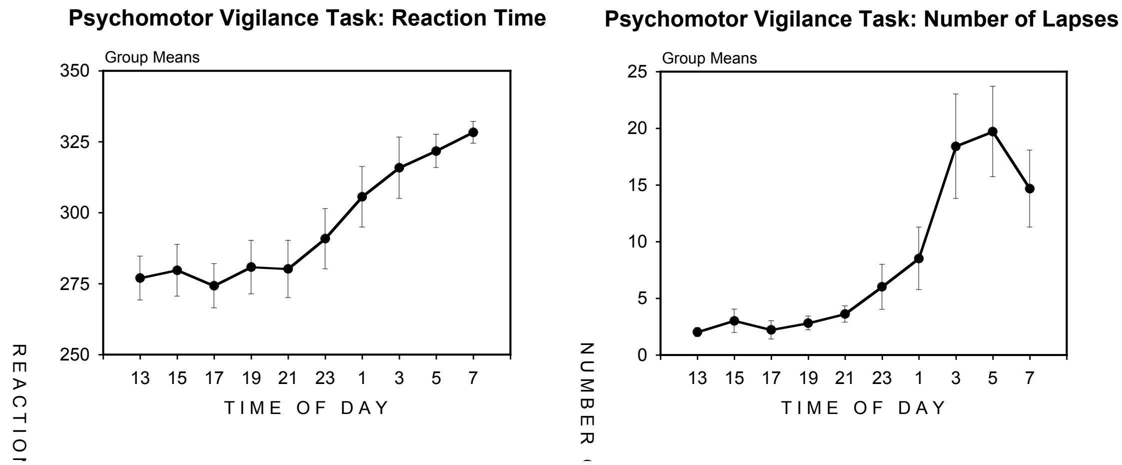

Figure 5: *PVT Performance: Speed (left) and Accuracy (right)*

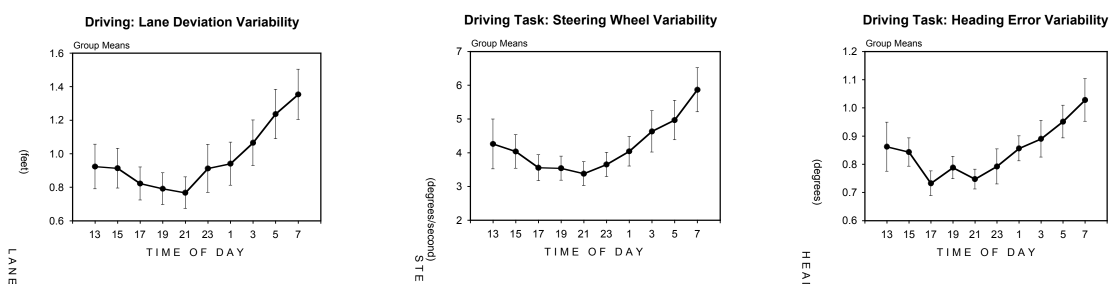

Figure 6: *Driving Performance: Measures for Lane Deviation (left), Steering Wheel (middle) and Heading Error (right)*

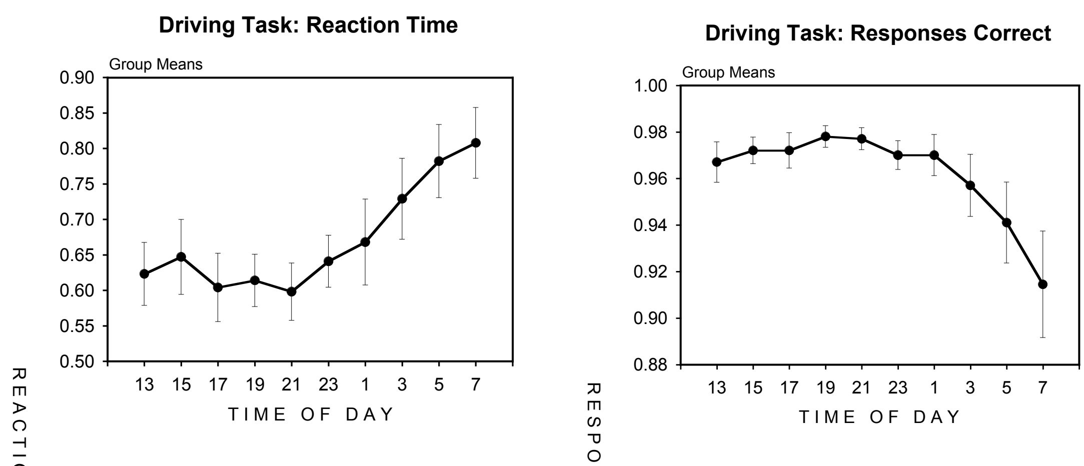

Figure 7: *Divided Attention Task During Driving: Speed (left) and Accuracy (right)*

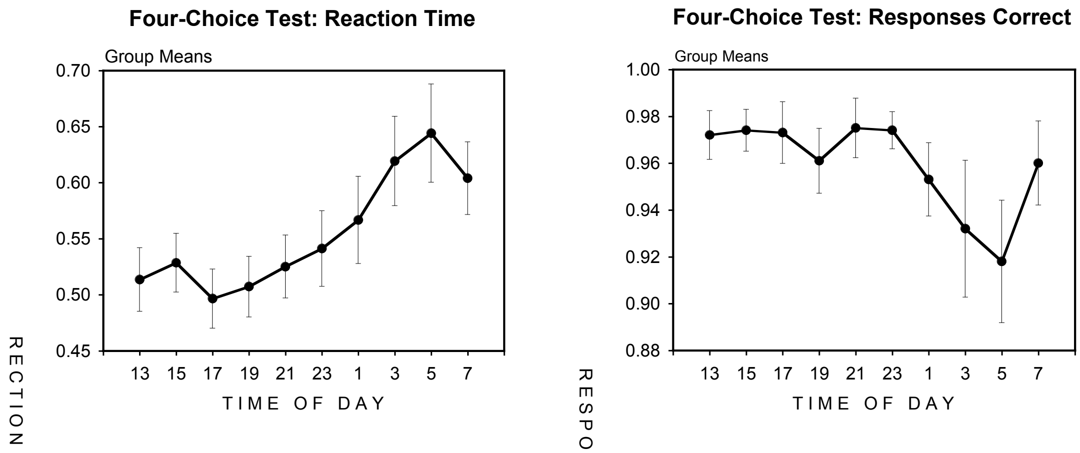

Figure 8: *Four-Choice Test: Speed (left) and Accuracy (right)*

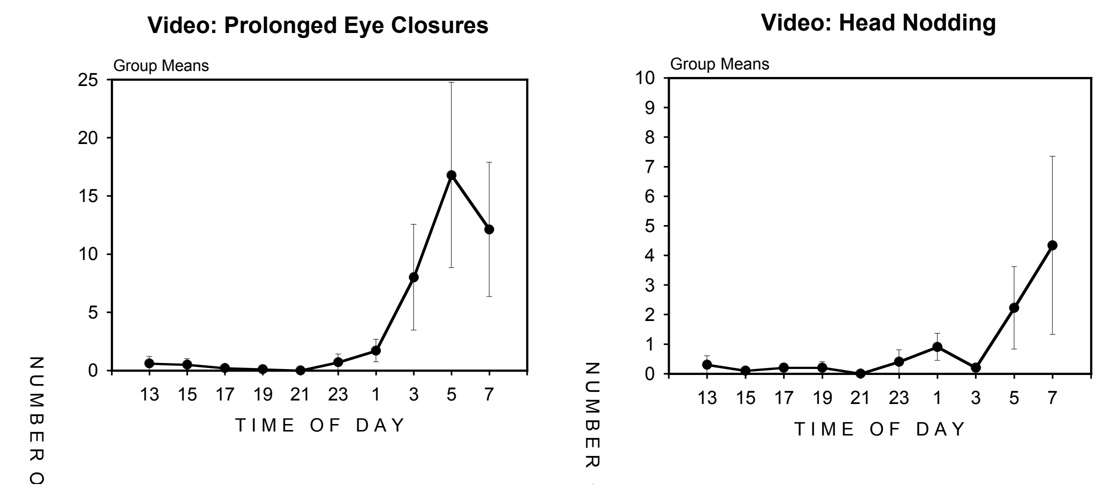

Figure 9: *Behavioral Signs of Sleepiness: Frequency of Prolonged Eye Closures (left) and Head Nodding (right)*

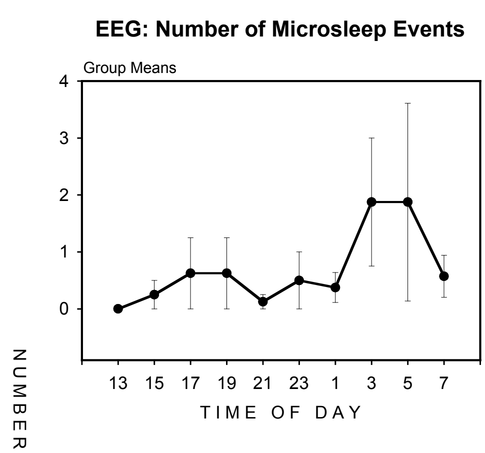

Figure 10: *Frequency of EEG Microsleep Events*

The Testbed parameters for subjective measures (alertness, mood, motivation, concentration, fatigue), the three performance tests (accuracy, speed), behavioral signs of sleepiness (frequency of head nodding and prolonged eye closures) and EEG microsleep events clearly showed the expected circadian trend, with consistently impaired alertness and performance levels at night. This confirms that the study protocol was appropriate for measuring a range of performance and alertness levels, which was needed for the correlation analysis.

Overall, the BLT Score showed a comparable trend, with most subjects having lower BLT scores in the early morning hours. However, the BLT results had a high inter- and intra-individual variability. Figure 11 shows the BLT group means across all test sessions (using the individual average for all BLT tests within the same test session). The pre- and post-driving averages for each test session are shown in Figure 12, with the post-driving average displaying a clearer circadian pattern. The results of the statistical analysis are described in section 3.1.3, below.

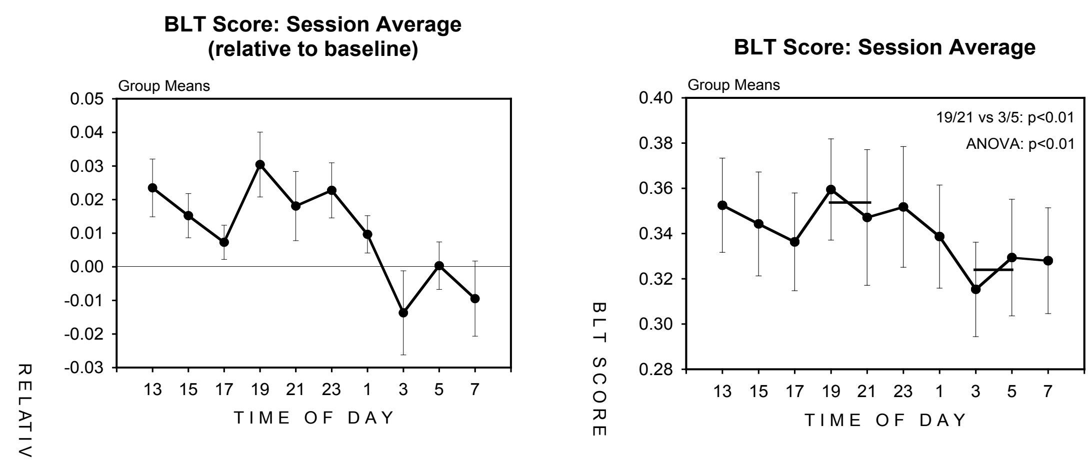

Figure 11: *BLT Score: Group Means for Relative (left) and Absolute (right)Values*

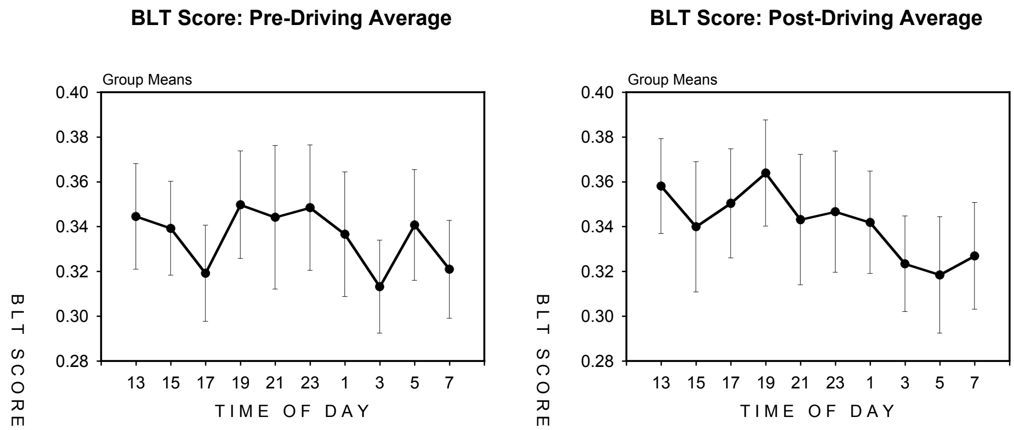

Figure 12: *BLT Score: Group Means for Pre- (left) and Post-Driving Average (right)*

1.1.3 Feasibility Testing (Aim 3)

Feasibility testing focused on validity and reliability. As indicated, a fitness-for-duty measure is considered valid if it detects some aspect of impairment, and is considered reliable when this detection is repeatable.

Validity Testing

Validity testing investigated the relationship between the BLT score and the Alertness Testbed measures. The main statistical method was correlation analysis. The feasibility criteria required statistically significant correlations for at least one alertness parameter and one performance parameter. On the group mean level, significant correlations were found between the BLT score (group means and post-driving averages) and almost all alertness and performance measures (Table 2). However, on an individual level, correlations were less strong and varied between subjects.

	BLT - Group Mean (Session Avg.)	BLT - Group Mean (Post -Driv. Avg.)
VAS - Alertness	0.774**	0.863**
VAS - Mood	0.764*	0.849**
VAS - Motivation	0.788**	0.866**
VAS - Concentration	0.725*	0.820**
VAS - Fatigue	0.742*	0.855**
Thayer - GA/DS	0.756*	0.847**
PVT - RT	-0.748*	-0.842**
PVT - # Lapses	-0.825**	-0.902***
Driving - LD Var.	-0.674*	-0.789**
Driving - SW Var.	-0.653*	-0.710*
Driving - HE Var.	-0.567	-0.686*
Driving - RT	-0.733*	-0.849**
Driving - % Corr ect	0.619	0.701*
Four-Ch. - RT	-0.781**	-0.920***
Four-Ch. - % Corr.	0.670*	0.717*
# Microsleep	-0.721*	-0.703*
Video - PEC	-0.704*	-0.847**
Video - HN	-0.434	-0.561

Table 2: Correlation Analysis: Correlation Coefficients and Significance Levels (* p<0.05, ** p<0.01, ***p< 0.001)

In addition to correlation analysis, statistical comparisons between daytime and nighttime windows were performed (t-statistics). If the BLT device was measuring some aspect of fatigue, it was expected to find significant differences between day and night. The daytime time window included the 1900 and 2100 session (a time period of high alertness which sleep physiologists refer to as "forbidden zone of sleep"), and the nighttime window included the 0300 and 0500 sessions (see Figure 11). The group BLT means differed significantly between the two time windows (p<0.01), and ANOVA detected a significant time-of-day effect (p<0.01). Similarly as with correlation analysis, within-individual analysis showed less strong results.

1.1.4 <u>Reliability Testing</u>

A fitness-for-duty device is considered reliable when the detection of impairment is repeatable in a test-retest situation. In this test protocol, subjects were required to take multiple BLT tests during each test session. The feasibility criteria required use of the pre- and post-driving BLT data across all test sessions as a basis for reliability testing. Correlations between the circadian time series of pre- and post-driving BLT results were performed. It was considered that the driving task itself may introduce sleepiness, and therefore create a difference between the pre-and post-driving BLT results. However, it was assumed that this potential difference would be a systematic one and therefore should not considerably affect the correlations between the two time series.

The analysis showed that the BLT score did not pick up a consistent difference between pre- and post-driving averages (which was seen, for example, in the subjective alertness measures). Correlations between pre- and post-driving BLT measures were not significant. The correlation for the group means resulted in r=0.514. Correlation analysis for individual subjects resulted in lower coefficients. In addition, correlations between BLT scores of paired tests (e.g., pre-driving test pair, post-driving test pair) were performed, but did not result in statistically significant results.

1.2 Conclusions

Alertness Testbed setup confirmed

The results of the study clearly showed the appropriateness of the Alertness Testbed setup. The Testbed measurements (subjective, alertness and performance measures) were clearly sensitive to circadian alertness and performance fluctuations and to impairment measurement.

Sleep deprivation protocol should be extended

As seen Figure 11, lower BLT scores (group means) were measured in sleep-deprived subjects in the early morning hours. Compared to the individual BLT baseline values (established during the hour before the start of the experimental protocol), this detected impairment was of moderate degree. It is concluded that the sleep deprivation protocol should be extended during Phase 2 testing in order to reach more extreme levels of sleep deprivation. This is particularly appropriate as the BLT technology is, by design, a fitness-for-duty device that is intended to detect extreme levels of sleepiness (employees working extreme work schedules and excessive overtime, or workers who experience extreme sleep deprivation for non-work reasons). Some subjects don't reach extreme impairment levels during one night of sleep deprivation, particularly if they enter the study without prior chronic sleep deprivation. Also, 24-hours of sleep deprivation are usually part of most standard shift work schedules (e.g., shift of a night shift block). Practically, rejections by a fitness-for-duty device should be an exception, not the norm, thus it is intended to detect more extreme levels of sleep deprivation than may have been achieved during the 24-hour sleep deprivation protocol used in Phase 1.

Validity of BLT score confirmed

The feasibility assessment supported strongly the validity of the BLT as an impairment measurement device. Overall, the BLT score correlated well with the Alertness Testbed measures, and it was shown that the BLT score detected nighttime impairment.

Reliability testing guides priorities for Phase 2 study

As expected, reliability of the BLT test was compromised by the random nature of the sequence of screens presented in a given BLT test. That is, the degree of difficulty varied considerably between screens, and the mix of easy and difficult screens within a given BLT test was not controlled. This limitation compromised the repeatability of test results in test-retest situations (e.g., subsequent tests in pre-driving and post-driving test pairs). Figure 13 illustrates the high test variability. The figure compares, for each individual, the standard deviation of BLT scores averages a) across all tests within the experimental protocol, b) across test pairs (two consecutive tests pre- and post-driving) and c) across the five tests of each test session. The highest standard deviations were found when averaging across all tests within the experimental protocol (as expected due to circadian fluctuations). However, the standard deviation for paired tests (test conducted in immediate sequence) was higher than the one for session averages. As the session averages include test before and after certain performance tests, the respective standard deviations should contain some variability due to the potentially fatiguing effects of the tests within the session or the whole session. On the other hand, for a test with optimal reliability, the standard deviation of the paired tests should be close to zero. The results indicate the limited reliability of the current BLT prototype.

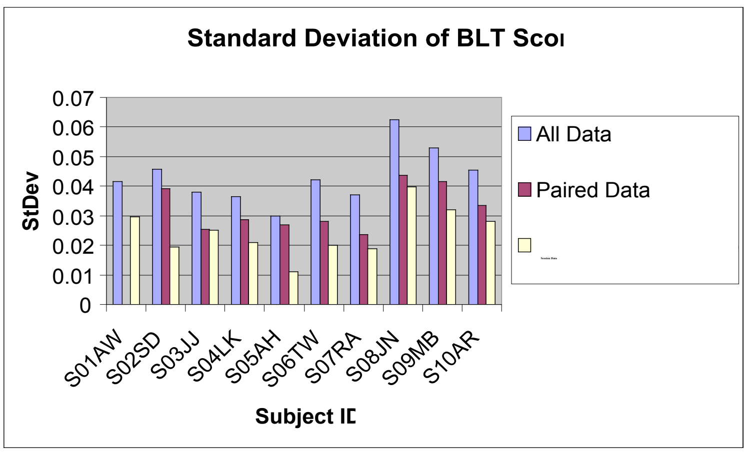

***Figure 13**: Variability of BLT Scores: Standard Deviations of BLT scores averaged across a) all test within the experimental protocol, b) across paired tests (two consecutive tests pre- and post-driving), and c) across the five tests of each test session.*

However, despite the reliability limitations, validity of the BLT score could still be demonstrated as we used BLT session means (average of the five BLT tests within each session) for statistical validity analysis.

The results of reliability testing guide the priorities for the Phase 2 study. The redesign of the BLT device will particularly focus on substantially increasing test stability.