Final Report

The Effects of Individualized Actigraph Feedback on Fatigue Management in Railroad Engineers

Patrick Sherry, Ph.D.
&
Karen E. Philbrick, Ph.D.

University of Denver

This research was sponsored in part by Grants from the
National Center for Intermodal Transportation,
The Federal Railroad Administration,
and the Burlington Northern Sante Fe Railroad

Executive Summary

A total of twenty-one participants completed the three-month study designed to assess the functionality of improving individual sleep habits with actigraph performance feedback. There were eleven participants in the feedback gro

Introduction

History of Fatigue in the Transportation Industry

The study of the role of fatigue and transportation has a long history. As early as 1907, Congress enacted the Hours of Service Act to enhance railroad safety by limiting the number of hours that railroad engineers and other railroad employees could work. However, the earliest published study of the effects of fatigue on locomotive engineers was in 1971 (Grant, 1971). Due to the ever-changing complexity of the demands faced by drivers and operators in all modes of transportation, the topic of fatigue continues to be the focus of intense study (Sherry

proneness" existed. This became a model for explaining and understanding safety thinking and research for almost 50 years (Cooper, 1998).

Heinrich (1931) however, proposed that accidents were caused either by an unsafe act, an unsafe condition, or both. His theory was termed Heinrich's *Domino Model of Accident Causation* and it brought in to play the idea that safe behavior was important as well as the roles that behavior, conditions, or the situation played. Essentially, the Domino model postulated that accidents were caused by a sequence of events, which covered five distinct phases. The first phase was considered the hereditary and environment of the person which would predispose them to act in a certain way. Heinrich argued that each of these was like a series of dominos arranged in such a way that if one fell then the others were likely to fall iny .5636 Tm(tha2.6j11

performance factors and argued for the focus on the overall management system, particularly in relation to the implementation of the organization's strategic decisions.

Haynes, Pine, & Fitch (1982) evaluated the effectiveness of an intervention package (feedback, competition, and incentives) in reducing the accident rate of urban transit operators. One hundred operators were divided into teams and offered rewards for accident-free driving over 18 weeks. Results showed a 24.9% reduction in accident rates, establishing a definite link between the intervention and reduction in accident rates, severity, and cost.

Karan and Kopelman (1987) provided outcome feedback regarding the actual frequency of accidents at a vehicle dispatch and maintenance facility. This outcome feedback was not provided at two similar comparison facilities. Over a 43-week experimental period, the rate of vehicular accidents declined by roughly 5% in the experimental facility, while accidents increased by roughly 17% in the two comparison facilities—thus, there was an overall improvement of approximately 22%. Concurrently, the rate of industrial accidents declined by roughly 12% in the experimental facility versus an increase of 4% in the comparison facilities--an overall improvement of approximately 16%.

There are numerous examples of the application of this type of model to the occupational safety arena. Two studies by Sulzer-Azaroff (1981, 1997) demonstrated the application of these principles to an industrial laboratory setting and a nursing home. Both situations met with considerable behavioral change.

A similar model of factors that affect safe work performance was suggested by Geller (1998). This model included the Person, Environment, and Behavior variables in a model labeled the Safety Triad (Geller, 1989). In this model, the three factors are thought to be dynamic and interactive, such that changes in one factor eventually impact the other two. For example, behaviors that reduce the probability of injury often involve environmental change and lead to attitudes consistent with safe work performance. According to Geller, the behavior based approach starts by identifying observable behaviors targeted for change and the environmental conditions and contingencies that can be manipulated to influence the target behaviors in desired directions.

A second type of behavioral influence on occupational safety is the role of antecedent factors such as psychological or attitudinal influences. Most traditional safety programs emphasize a need to increase employee's awareness of safety hazards and in so doing prevent injury. This is considered an environmental influence on behavior. The influences on behavior may

Writing about the need to improve the environmental conditions under which behavior change might be maintained, Krause, Hidley and Hodson (1990) promoted the idea that a safety corporate environment needed to be created so as to sustain the behaviors that needed to be changed.

external consequences seemed to be associated with undesirable spread of presumed counter control effects.

This review then, has identified the behavioral based safety approach as one that may be useful in addressing the concerns that operators have regarding the adoption and utilization of various OBSM systems. Perhaps, through the use of behaviorally based safety concepts, the resistance and reluctance to engage in safe work behaviors that involve the utilization of OBSM systems can be reduced.

Effects of Feedback

Much of the success of the behavioral approach to safety is based on the notion that feedback of any type can have a positive effect on safety behavior. The basic idea comes from operant theory (Skinner, 1947) as well as cognitive – behavioral theories on behavior change (Beck, 1993). However, the effects of feedback on performance have only received attention in the psychological research literature.

A review article by Balzer (1989) found that in some conditions feedback interventions improve performance, in other conditions feedback interventions have no apparent effect on performance, and in yet others feedback interventions debilitate performance. These conditions or moderators of the effect of feedback interventions (FIs) on performance are poorly understood and go far beyond the view that feedback interventions improve performance unless the feedback is too negative. However, many researchers still assume that feedback interventions consistently improve performance.

Two meta-analyses, testing theories that contained feedback as a component, found only a weak contribution of feedback to performance. First, Harris and Rosenthal (1985) tested several hypotheses designed to explain the well-documented beneficial effect of expectations of others (agents) on one's performance. When agents (primarily teachers) expect high performance from others (primarily students), they may provide more feedback, more challenging goals, and create a better climate for the students. This meta-analysis showed that the amount of feedback provided by the agent had only a meager effect on performance ($r = .07$),

FI cues that seem to direct attention to task-motivation or task-learning processes may augment FI effects on performance. This pattern of findings provides reasonable support for the first two propositions. Specifically, both praise and FI designed to discourage were postulated to increase attention to meta-task processes and were found to attenuate FI effects. Furthermore, both the attenuating effect of praise and the non-significant effect of an FI are not easily predicted by most FI-related theories. The debilitating effects of praise on performance received some direct experimental support both in the laboratory and in the field and were explained, respectively, by a model of self-attention (Baumeister et al., 1990) and by control the

Description of the Project and Data Collection Procedures

The current project was designed to obtain individual participation in the monitoring of fatigue through the use of individual fatigue monitors. The goal of this study was to determine whether individual feedback devices, such as actigraphs, could be useful for helping railroad employees better plan their sleep and wake activity. Project participation consisted of the completion of a consent-form, several survey questionnaires, a daily sleep log, and wearing an Actigraph, which measured sleep and work during the course of the project.

Prior to soliciting participation, it was necessary to identify those conditions that would exclude an engineer from participation. Specifically, those persons who were not able to wear the activity monitors for the full 60-day period were not eligible to participate. Similarly, persons who had a diagnosed condition that affected their sleep patterns, and persons who were working a schedule that would be dramatically different from the typical pool assignment also were not eligible for participation.

Once an individual agreed to participate, he or she was notified that data collected as a result of participation in the project would only be shared with the participants themselves. They were also notified that the BNSF agreed

	Engineer	Conductor	Trainmaster	Management
Chicago	12			
Creston	10	1		
LaCrosse	3			

Galesburg

The Galesburg Time 2 Fatigue Survey was given June 15, 2002. While many of the content questions were intentionally identical to the Time 1 survey (to aid in making comparisons), the demographic data collected was slightly different. Specifically, gender, race, age, and educational attainment were not asked. Fifty-nine individuals completed the Time 2 survey. Of these respondents, 22 were conductors, 35 were engineers, and 2 did not indicate their craft.

The Galesburg Time 3 Fatigue Survey was given July 15th, 2002. This survey was given only to those individuals who were also wearing an actigraph. This differs from the Time 1 and 2 survey collection procedure, as these were given to all engineers and conductors at the depot

At the end of the second month of the study, each participant received a $25.00 gift certificate to a local restaurant. Similarly, at the end of the third month each participant received an additional $25.00 gift certificate, for a combined total of $50.00 for wearing an actigraph for three months.

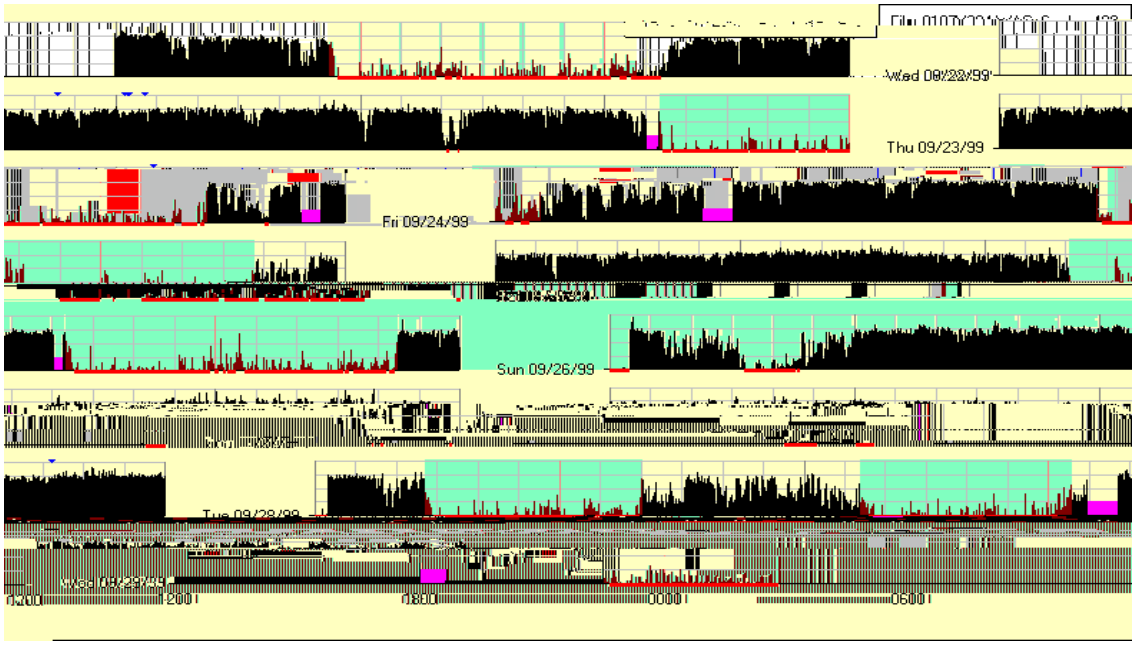
A total of three measurements took place in 2002. Again, researchers arrived on site in Galesburg Illinois on May 15th, July 15th, and on August 15th. At the Time 1 measurement, fatigue surveys were administered and actigraphs were distributed. At the Time 2 measurement, fatigue surveys were again administered and a brief meeting was held with each actigraph participant. During this meeting, half of the participants received a feedback actigragh and the batigu1tssw52 31re aga

Additional Information Regarding Study Materials

Actigraphs -- These devices are essentially motion detectors. They are able to keep track of the amount of body movement that occurs. They are mechanical and do not harm the individual wearing them. They do not keep track of pulse or electrical activity. They must be worn continuously but should be taken off for showering or bathing or vigorous exercise. Various studies over the years have demonstrated a very strong relationship between body movement and sleep.

Here is what an Actigraph looks like….

Participants were asked to wear the device for 30 days. At the end of the thirty-day periody



Researchers were on-site to address an

Self-Report Survey Results

There are several ways to determine whether a person is fatigued or not. We can simply ask the person if they are fatigued or sleepy or tired. We can examine their brain waves, we can examine their performance, or we can see how long it takes them to fall asleep. All of these approaches have pros and cons. In field settings, like the railroad, it is most economical to ask participants to complete standardized questionnaires that have been correlated with laboratory findings. This technique is typically used to make preliminary assessments of persons who are presenting with possible sleep disorders in medical settings. These questionnaires then give a reasonable indication of the level of fatigue and tiredness that persons are experiencing.

Comparisons Between Time 1, Time 2, and Time 3

To determine whether scores on any of the self-report indices changed significantly for the actigraph participants from Time 1 to Time 3, one-way ANOVA's were conducted. Significance levels, in addition to means and standard deviations, are reported below.

	N	Mean	Standard Deviation	Significance (2-tailed)
SSS-Current Alertness	38	3.3158	1.4539	.590
SSS-Alertness Last Week	38	4.0263	1.1505	.898
Eppworth Sleepiness Scale	40	8.6000	4.0115	.396
Denver Job Satisfaction Scale	41	2.4797	.9281	.000
Denver Fatigue Adjective Checklist	41	2.4111	.7923	.441
Denver Sleepiness Scale	41	2.9106	.5376	.728
Denver Depression Scale	41	2.9512	6.8042	.361
Denver Anxiety Scale	41	2.1744	.6895	.384
Denver Stress Scale	41	2.7073	.7069	.695
Denver Quality of Life Scale	41	2.4439	.7906	.265
Shift Work Index - Exhaustion	41	2.7195	.8124	.256
Shift Work Index – Depression	41	2.9824	.5937	.001
Shift Work Index - Quality of Life	41	2.6301	.9955	.000

Results of these analyses yielded three significant findings. Specifically, there are significant differences on the Denver Job Satisfaction Scale between the different points of measurement, (F (2,38) = 25.805, p<.000). These results suggest that satisfaction with one's job increased over the course of this study. In fact, at the Time 3 measurement, 38.9% of respondents indicated that they were satisfied with their job, to either a "Considerable" or a "Very Great Degree".

Results of the one-way ANOVA on the Shift Work Index – Depression scale were also significant, (F (2,38) = 9.330, p<.001). This finding indicates that, from Time 1 to Time 3, subjective feelings of depression decreased, as evidenced by the endorsement of fewer items assessing unhappiness and lethargy, while feelings of well-being increased. Participants felt more capable of making decisions (mean = 3.89) and experienced increased enjoyment of day-to-day activities (mean = 2.78).

The Shift Work Index – Quality of Life scale yielded significant results as well (F (2,38) = 10.14, p<.000). Compared to measurement Time 1 and Time 2, participants felt that they were making more attempts to get rest and that sleep/rest patterns had changed in a positive way. Combined, changes such as these had a positive impact on the quality of life of the engineers in this study.

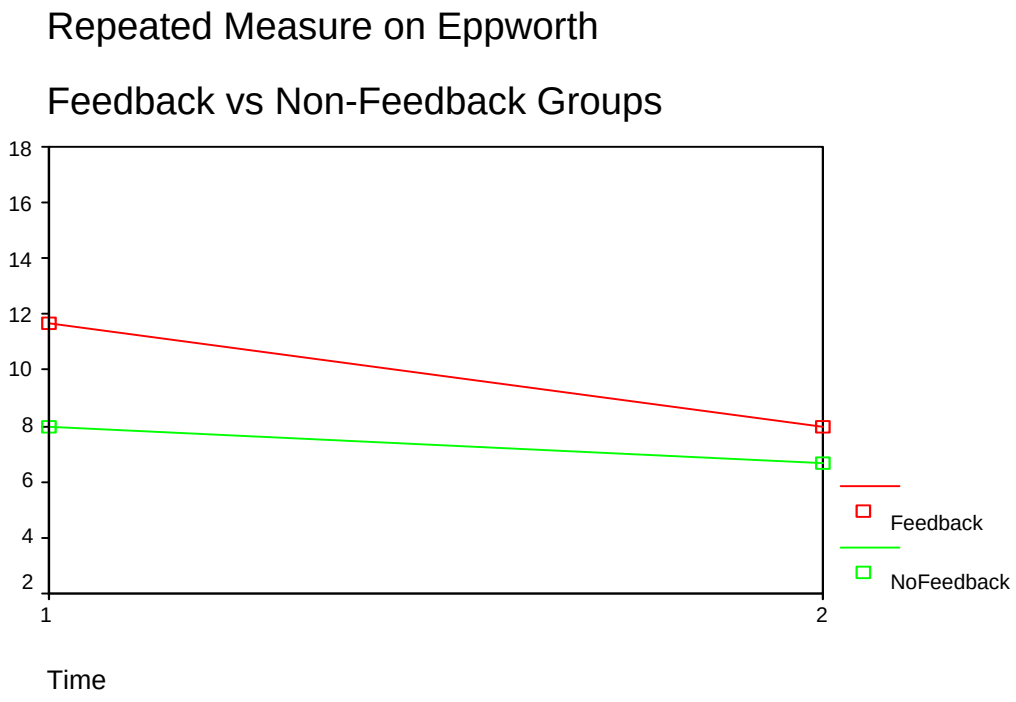
Comparison of Feedback vs. Non-Feedback Groups on Self-Report Measures

To understand the effects of individualized actigraph feedback on fatigue management in railroad engineers, a variety of measures were employed and it was necessary to compare the performance versus the non-performance actigraph participants on these measures. Following the discussion of significant results, the table below presents the means, standard deviations, and significance values for all indices.

The analysis of variance statistical technique was used to test for significant differences between the feedback and non-feedback groups. One of these analyses yielded significant results. Specifically, significance was found for the Denver Depression Scale, (F (1,16) = 4.73, p<.045). These results suggest that the non-feedback group reportedly experiences more symptoms of depression as compared to participants in the experimental group. Engineers in the feedback group report enjoyment of daily activities and feeling reasonably happy. However, although this finding is significant, the mean score on this scale for the feedback group is 1.70 on a 5-point scale. This suggests that while there is a notable difference between the two groups, all participants may suffer from some depressive symptoms. Often times this can be due to excessive demands at work and home and may suggest that these participants may benefit from adopting more adaptive coping skills.

of Life Scale, the Shift Work Index – Exhaustion Scale, the Shift Work – Depression Scale, and finally the Shift Work Quality of Life Scale. This statistical analysis was chosen since within-subjects variables always involve taking repeated measurements from each subject, as was done for these groups of participants at Time 1, Time 2, and Time 3 measurements. In within-subject designs, the same subjects are tested in each condition. Therefore, differences among subjects can be measured and separated from error with this method of analysis.

While the overall results were not significant for these analyses, some interesting trends were noted. Below are graphs representing the trends that were highlighted via the use of repeated measures analysis of variance.



This graph shows the results for the means plot repeated measures analysis of variance for the Eppworth Sleepiness Scale. Results indicate that there were large differences between the feedback and non-feedback groups at pre-test. These differences became much smaller at post- test. While the overall results are not significant, there is clearly a trend showing that the feedback group has improved, indicating that the likelihood of the feedback participants "Dozing or Falling Asleep" in eight different types of situations is reduced. The slope of the feedback line is much steeper than that of the non-feedback group. Clearly, larger sample sizes would permit a more robust test of the effects of performance feedback on the Eppworth Sleepiness Scale.

In terms of the sleep watch helping feedback participants to manage their fatigue, findings were significant (F (1,14) = 5.24, p<.038). When asked to what degree the watch made the users more aware of their need for rest and sleep, the mean response was 3.90, thus indicating that the watch made them more aware of their need for rest/sleep to a "Considerable Degree". In fact, 70% of engineers in the experimental group indicated that the performance readings made them more aware of their fatigue levels to a "Considerable or Very Great Degree". Thus, it appears that the performance readings helped people manage their fatigue because they could base decisions on when to rest and when to complete domestic chores as well as activities of daily living based on percent performance.

Participants became significantly more aware of their fatigue levels as a result of percent performance feedback (

Comparison of Feedback vs Non-Feedback Groups

10	2.6000	1.0750	.679
8	2.3750	1.1877	
18	2.5000	1.0981	
9	3.5556	1.6667	.038
7	1.7143	1.4960	
16	2.7500	1.8074	
9	4.4444	1.1304	.002
7	1.7143	1.7043	
16	3.2500	1.9494	
10	3.9000	1.1972	.125
7	3.0000	1.0000	
17	3.5294	1.1789	
10	3.9000	1.2867	.019
8	2.2500	1.3887	
18	3.1667	1.5435	
10	2.9000	1.5239	.168
8	1.8750	1.4577	
18	2.4444	1.5424	
10	2.6000	1.5776	.332
8	1.8750	1.4577	
18	2.2778	1.5265	
10			

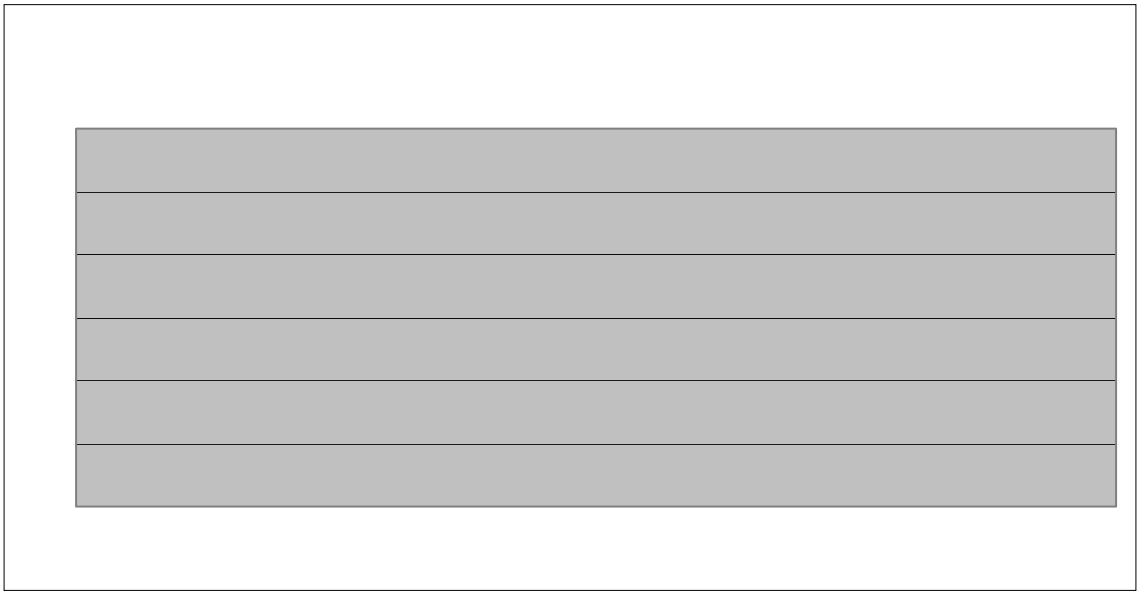
Questionnaire Results for All Survey Respondents

are not at peak performance when at work and that these same individuals may be in danger of developing a serious sleep debt.

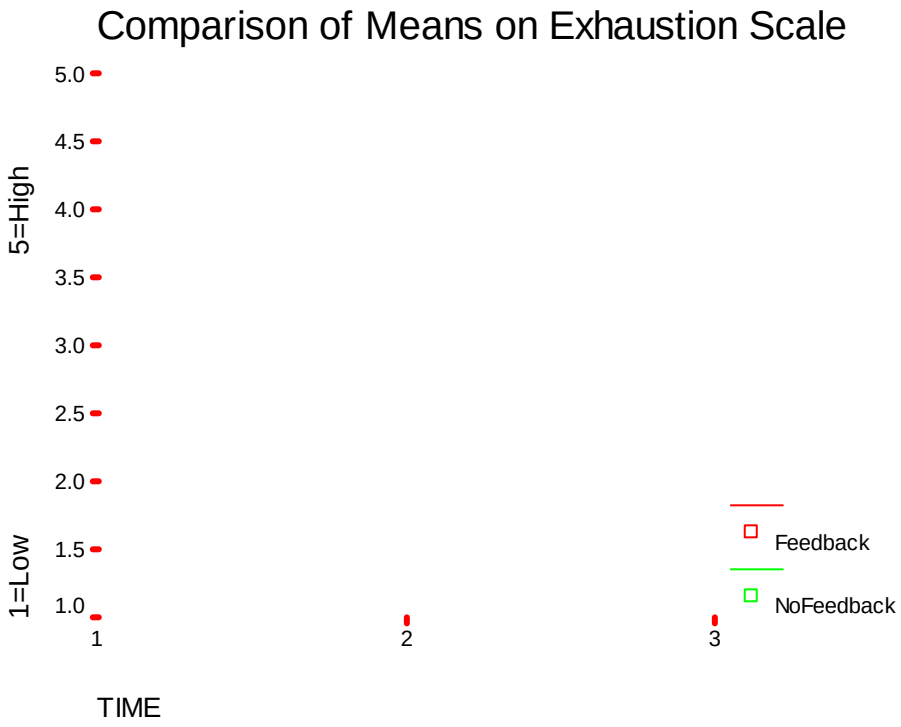
The other instrument that is commonly used in assessing sleep and fatigue related difficulties, is the Eppworth Sleepiness Scale. The Eppworth Scale is routinely used in Sleep Disorders Clinics around the world to identify persons at risk for sleep apnea and other sleep disorders. Scores on this instrument are indicative of a high potential for sleep disorders.

The Eppworth Sleepiness Scale assesses the degree to which individuals are likely to doze off or fall asleep in eight different types of situations. In general, the cutoff of 8 or 9 points is thought to indicate some risk. Scores over 10 or 12 are thought to be in the clinically significant range meaning that additional study is needed. For the purposes of this study then we would expect that persons above 10 would be experiencing very high and abnormal degrees of tiredness in comparison to the normal population.

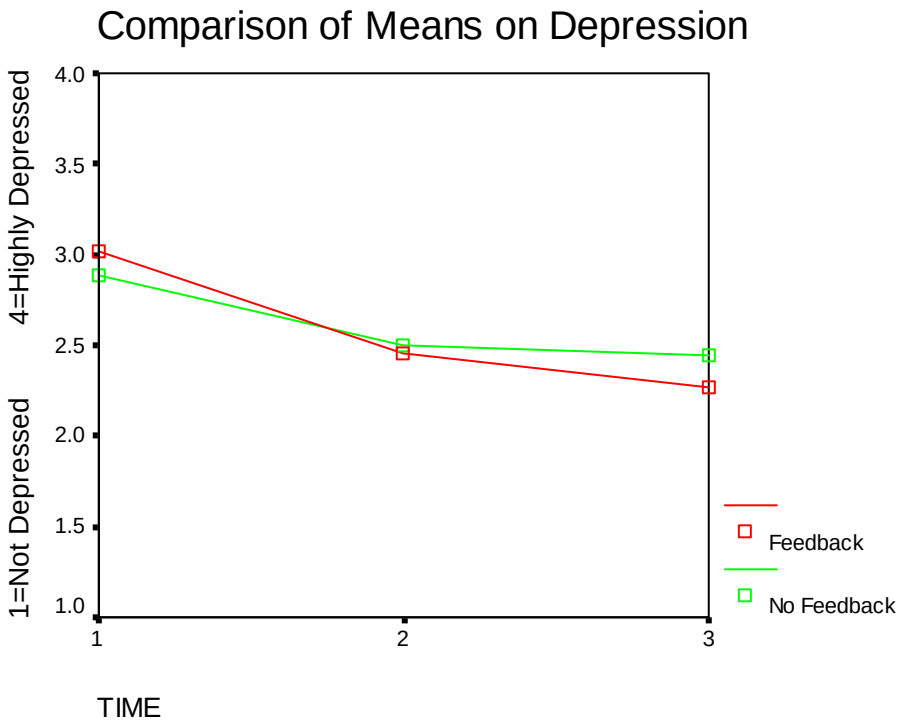
Eppworth Sleepiness Scale



A similar result was found with the Exhaustion scale which measures the intensity with which a number of fatigue related adjectives are endorsed. As can be seen from the graph below there is an overall tendency for the scores to slope to the right which indicates a gradual improvement in fatigue as measured by the exhaustion scale.. In addition, the two groups differ in overall magnitude of exhaustion. Thus, there is a nearly significant difference between the experimental and control groups such that the feedback group is somewhat higher than the non-feedback group. Overall, there is a significant reduction in exhaustion over time.



There were however nearly significant differences between the two groups at post-testing on depression, suggesting an interaction between the feedback condition and mood. As can be seen from the graph, the feedback group is *lower* on depression at Time 3 than the non-feedback group.



Discussion

Overall, study participants in the experimental group reported that the performance actigraph was a useful tool for fatigue management. While there were not statistically significant differences between the feedback and non-feedback actigraph participants on indices assessing stress, fatigue, or anxiety, significant differences were found between the two groups on questions assessing the usefulness of the actigraph. Specifically, the most robust finding indicated that the performance sleep watch helped participants in the experim

Watson, J.B. (1930). Behaviorism. New York: Norton.

Weller, D., & Sherry, P. (1992). Role of supervisor support in buffering the