Human Factors: The Journal of the Human Factors and Ergonomics Society

http://hfs.sagepub.com/

Long Workdays versus Restdays: Assessing Fatigue and Alertness with a Portable Performance Battery

Roger R. Rosa and Michael J. Colligan

Human Factors: The Journal of the Human Factors and Ergonomics Society 1988 30: 305
DOI: 10.1177/001872088803000305

HUMAN FACTORS, 1988, 30(3), 305–317

Long Workdays versus Restdays: Assessing Fatigue and Alertness with a Portable Performance Battery

ROGER R. ROSA[1] *and* MICHAEL J. COLLIGAN, *National Institute for Occupational Safety and Health, Division of Biomedical and Behavioral Science, Cincinnati, Ohio*

A test battery designed to assess psychological/behavioral fatigue was used to compare restdays to a workweek of five 12-hr days at a data entry job simulation. Across both workdays and restdays, the battery was presented at regular intervals to test for fatigue effects and diurnal variations. Increased data entry errors across the workday and the workweek as well as subjective reports of increased tiredness on workdays indicated that the work regimen was fatiguing. Test battery performance paralleled those results. On workdays, as compared with restdays, grammatical reasoning was faster but less accurate; digit addition was slower; simple, dual, and choice reaction times were slower; and hand steadiness decreased. The results demonstrated the sensitivity of the battery to long hours of work. The results are discussed in terms of work-rest differences, changes across the workweek, diurnal variations, and cognitive demand.

INTRODUCTION

There is a considerable body of literature suggesting that shiftwork in general, and nightwork in particular, may have adverse consequences for the health and well-being of the individual worker (e.g., Johnson, Tepas, Colquhoun, and Colligan, 1981; Rutenfranz, Colquhoun, Knauth, and Ghata, 1977). For the most part these problems have been attributed to the fact that shiftwork requires individuals to work at times that are at variance with their customary physiological, psychological, and social circadian rhythms (Aschoff, 1981; Wever, 1981). Not surprising, then, is the growing interest in exploring the feasibility of alternate work schedules. One approach that has received attention recently is the 12-hours-per-day compressed workweek. Although there are a number of variations (see Colligan and Tepas, 1986), the underlying strategy is to shorten the workweek to three or four days by increasing the length of the work shift to 12 hours. The presumed advantage is that the shiftworker has more nonworking days per week, thereby gaining larger blocks of time for recuperation and leisure. The potential risk, on the other hand, is that the additional four hours per shift may

[1] Requests for reprints should be sent to Roger R. Rosa, National Institute for Occupational Safety and Health, 4676 Columbia Pkwy., Cincinnati, OH 45226.

result in fatique-induced decrements in performance and increased accident susceptibility.

Our review of the literature on compressed workweeks (Rosa and Colligan, 1986; Rosa, Wheeler, Warm, and Colligan, 1985) revealed persistent concerns about feelings of increased fatigue associated with long workdays. These concerns have made some organizations hesitant to change from the conventional five-day/40-hr week to some form of compressed week requiring longer workdays. They hesitate despite employees' overall satisfaction with the compressed week and willingness to tolerate or adapt to increased subjective fatigue. Concerns about increased fatigue are supported by both laboratory (Colquhoun, Blake, and Edwards, 1968, 1969; Rosa et al., 1985) and work site (Volle, Brisson, Perusse, Tanaka, and Doyon, 1979) comparisons of 8-hr and 12-hr days. These concerns pointed to a need for systematic study of long workdays, preferably at the work site. Our own efforts in this area began with laboratory development of a methodology for assessing performance and subjective fatigue and alertness associated with long workdays. We have recently extended these efforts to work site evaluations of 12-hr/day schedules (Rosa, Colligan, and Lewis, 1986).

Our methodology involves a test battery containing measures for a range of psychological functions. A detailed rationale and complete description of the battery can be found in Rosa and Colligan (1986) and Rosa et al. (1985). In brief, the battery was designed for work site testing in a manner that minimizes interference with the worker's job responsibilities. As such, it is portable, brief, and easy to administer. In addition, specific tasks in the battery require minimal training of the subject. The built-in flexibility of the battery makes it easy to select any subset of the tests for a particular application and to time any task to minimize interference with the worker's job.

Our first laboratory study (Rosa et al., 1985) demonstrated greater decrements in performance, increased subjective fatigue, and decreased alertness in a 12-hr, four-day week as compared with an 8-hr, six-day week. That study suggested that the battery was sensitive to differences in work scheduling and could be used in field studies. However, because no resting condition was included in that study, the extent to which the test battery was sensitive to work fatigue was unclear. Therefore, we compared 12-hr workdays to resting days of leisure activities in the present investigation. Our study design tested the limits by comparing restdays with a somewhat atypical work schedule of five 12-hr days. Except for occasional overtime situations, such a schedule would normally not occur in organizations using the extended workday. In the present study, however, it served the purpose of exposing the subjects to a fairly rigorous work regimen in order to increase the chances of finding work-rest differences on the test battery. Presentation of the battery at regular intervals across these workdays and restdays allowed for potential diurnal variations.

METHOD

Subjects

Eleven male volunteers were recruited from a temporary clerical/office employment service for participation in the study. In exchange for their participation, they were paid \$5.00 an hour for the first 40 hours each week and \$7.50 for every hour exceeding that, plus a bonus of 50 cents per hour if they completed the entire study. Subjects were selected according to the following criteria: (1) ages between 18 and 25 years, (2) no major medical or psychiatric disorders, (3) normal sleep patterns (i.e., 7–9 hrs per night), (4) no history of shiftwork within the past six months, and (5) normal (20/30 or better) corrected vision. Nine of the 11 subjects completed the study.

Simulated Work and Fatigue Test Battery Measures

Job performance was measured with a data entry task operated throughout the day. This task was periodically interrupted to complete the fatigue test battery. Included in the battery are a number of brief performance tasks and scales of subjective alertness and subjective fatigue. The performance tasks were designed to evaluate a range of psychological functions including cognitive abilities, perceptual-motor functions, motor skills, and sensory acuity. The tasks are briefly described in the following paragraphs. More background and detail can be found in Rosa and Colligan (1986) and Rosa et al. (1985). Both the work simulation and the fatigue test battery were presented on a microcomputer that scored all tasks and scales and stored the data.

Work Simulation: Data Entry Task

The data entry task was a simulation of work performed in any number of clerical jobs. Randomly constructed five-digit numerical strings were presented on the computer viewing screen at a rate of 20 per min. The subject was instructed to reenter each string on the computer keyboard as rapidly as possible. The score for the task was the number of data entry errors per 60-min work period. Errors included miscopying one or more digits in the string and failures to complete the string.

Fatigue Test Battery: Cognitive Abilities

Grammatical reasoning. This is a 4-min test of verbal analytical abilities coupled with a short-term memory requirement. It is a variation of a task devised by Baddeley (1968). A two-, four-, or six-letter stimulus string (e.g., JLNP) was initially presented for 2 s, then removed, and followed after 2 s with a conditional statement such as: J DOES NOT PRECEDE N. The subject then decided whether the statement satisfied the letter string by pressing pushbutton switches labeled TRUE or FALSE as quickly as possible. To minimize acoustic confusability, only the letters J, L, N, P, Q, and R were used. These letter strings were generated randomly for each trial. Sixteen conditional statements were presented during the task. They represented all permutations of the following dichotomies: precedes or follows, active or passive voice, positive or negative, and true or false. The order of presentation of the statements was also randomized. Scores for the task were the mean of the response latencies and the total number of errors for the 4-min test.

The entire test was conducted with a memory load of two letters, four letters, and six letters at each testing time. Three memory loads were included to test whether performance on tasks with higher cognitive demand peaked earlier in the day (Monk and Embrey, 1981).

Digit addition. This 3-min task was adapted from Williams and Lubin (1967). The subject was required to add two digits presented on the screen and to type the last digit of the sum on the keyboard. In a separate 3-min segment, the subject was required to add a constant to the sum of the two digits and to type the last digit of the final sum on the keyboard. The subject was requested to work as quickly as possible. Scores for the task were the total number correctly completed and the percentage of errors per 3-min period. The percentage of errors was transformed to an arcsine to better approximate the normal distribution in data analyses (Myers, 1979).

Two versions of the task were used so that differences based on the number of cognitive operations (two additions versus one) could be compared while keeping the stimulus input (two digits) and the response output (a single keystroke) the same.

Time estimation. Subjects were asked to produce two 15-s intervals by using a key press as a start and a stop signal. They were requested to count to themselves during the interval. Each time estimate was converted to a ratio of the estimated time divided by the target interval. The score for a given session was the mean of these ratios.

Fatigue Test Battery: Perceptual-Motor Functions

Simple auditory reaction time. Stimuli were sixty 43-ms bursts of white noise generated by a National Semiconductor M5807 white noise chip and presented via headphones at 72 dB (SPL) for 4 min. The mean interstimulus interval was 4 s with a variation of ±2 s. The subject's task was to respond as quickly as possible to each auditory stimulus. During the task, the subject maintained an upright sitting position with the forefinger of his or her nonpreferred hand resting on a pushbutton switch mounted in a response box and resting on a table. Scores for the task were the mean reaction time for the 4 min and the percentage of responses >1.5 s. These unusually long (or missed) responses are an indication of attentional lapses (gaps or blocks—see Bills, 1931). The percentage of gaps was transformed to an arcsine to better approximate the normal distribution in data analyses (Myers, 1979).

Choice reaction time. Stimuli were 100 random presentations of the words TRUE or FALSE presented on the viewing screen for 10 min. The subject's task was to respond as quickly as possible to each stimulus by pressing the appropriately matched TRUE or FALSE response button. During the task, the subject rested the forefinger and middle finger of his or her preferred hand on the TRUE and FALSE buttons, respectively. Other aspects of the task were the same as simple auditory reaction time. The task was divided into four periods, each with 25 signals (approximately 2.5 min per period), so that potential decrements with time on task could be observed. Scores for the task were the mean reaction time, the percentage of choice errors, and the percentage of gaps or blocks for each period. The percentages of errors and gaps were transformed to arcsines to better approximate the normal distribution in data analyses.

Dual-task performance: Grammatical reasoning and simple reaction time. As an added demand on processing capacity, the four-letter version of the grammatical reasoning task and the simple auditory reaction time task were performed simultaneously. Bursts of white noise were presented on 75% of the grammatical reasoning trials in one of six randomly selected times, prior to the presentation of the conditional statement. In order to minimize response interference, no auditory probes were presented after the conditional statement. Thus it was assumed that any potential interference attributed to dual-task performance occurred during the processing of the letter string in working memory. Responses to the reasoning task were made with the preferred hand and responses to the reaction time task were made with the nonpreferred hand. Subjects were instructed that their primary task was to maintain their best performance on the reasoning task while performing the reaction time task to the best of their ability. All other aspects of the tasks were the same as those tasks performed individually.

Fatigue Test Battery: Sensory Acuity

Two-point auditory discrimination. This task is a measure of the temporal resolving power of the auditory system; it was developed as an auditory counterpart to critical flicker fusion. Several pairs of 108-ms white noise pulses were presented via headphones at 72 dB (SPL) with varying periods of silence between them. Subjects then decided whether the stimulus pairs sounded like one continuous burst of noise or like two bursts of noise, and indicated their judgment by pressing the numeral 1 or 2 on the keyboard. Fusion threshold was determined by the dual-staircase method of limits (Dember and Warm, 1979). The threshold score was the length (in ms) of the silent gap between the two noise pulses. Discrete pairs of noise pulses were used instead of a continuous on-off series of noise pulses (the analogue of a flickering light) in order to avoid the confusion created by the perception of pitch at the modulating (on-off) frequency of a series of noise pulses.

Fatigue Test Battery: Motor Skills

Response alternation performance (tapping). This task required 3 min of repeated alternating taps on two keys with the middle finger and forefinger of the preferred hand. It was scored for the total number of taps and for the number of taps with an intertapping interval greater than 750 ms (lapses). The lapse value was based on pilot work with unfatigued subjects and represents the slowest tenth percentile of the total number of taps. The task was adapted from Naitoh (1981).

Hand steadiness. This task required centering a conductive metal wand inside a circular metal opening ⅛ inch (0.32 cm) in diameter for a period of 2 min. During the task no support was permitted for the hand, arm, or elbow. Touching the perimeter of the opening completed a circuit and activated a counter that sampled at a rate of 150 times per s. Score for the task was the percentage of time off-target. These scores were transformed to arcsines to better approximate the normal distribution in data analyses.

Fatigure Test Battery: Self-Report Measures

Stanford sleepiness scale. The seven-point Likert-type scale devised by Hoddes, Zarcone, Smythe, Phillips, and Dement (1973) has descriptors ranging from *very alert* to *very sleepy*. Subjects were instructed to choose the set of descriptors that best described their feeling of sleepiness at the given moment.

NPRU adjective checklist. Twenty-nine adjectives describing negative or positive feelings were presented in a randomized sequence and scored on a four-point intensity scale ranging from 1, *not at all*, to 4, *extremely*. Responses to adjectives describing negative feelings were summed for the Negative score and responses to adjectives describing positive feelings were summed for the Positive score (Johnson and Naitoh, 1974).

Fatigue Test Battery: Physiological Measure

Oral temperature. Sublingual temperature was taken at each administration of the test battery. The computer prompted the subjects to insert the thermometer, which was then held in the mouth for 5 min during completion of the self-report scales. Subjects then entered their temperature in response to another computer prompt. Tempa-DOT brand single-use disposable thermometers (Info-Chem, Inc., Fairfield, NJ) were used. Temperature was measured as an index of the circadian rhythm of physiological activation (Aschoff, Geidke, Poppel, and Wever, 1972).

Design

The subjects first worked a week of five 12-hr days (0700–1900 hrs) at the data entry job simulation. On each day they broke from the task at 0800, 1000, 1300, 1500, and 1730 hrs for testing on the fatigue test battery. Data from this workweek were considered practice/adaptation and were excluded from further analysis. After two days of home rest, the subjects returned to the laboratory. They were again administered the fatigue test battery at the designated times in the absence of any intervening work at the data entry task (first restday). They then worked the experimental week of five 12-hr days at the data entry job simulation with testing on the fatigue test battery at the times noted earlier. This workweek was followed by two days of home rest followed by a final laboratory restday of testing on the fatigue test battery at the designated times.

Procedure

Subjects were divided into two groups, as only six could be accommodated at one time. Five of six subjects completed the study in the first run and four of five subjects completed the study in the second run.

Seven days before the first workweek, each group reported to the laboratory for an introduction to the procedures. Each subject was assigned to an individual work cubicle designed to reduce visual contact among the participants. Each workstation contained a microcomputer and accompanying accessories (e.g., headsets for auditory discrimination testing, peripheral response buttons for reaction times) for the performance testing. On the introductory day, the data entry task was performed for 75 min and the fatigue test battery (45 min) was performed twice. The order of tasks in the fatigue test battery was the same for each subject and was maintained throughout the study.

Laboratory workdays began one week after the introductory session. On workdays the subjects maintained the schedule shown in Table 1. Talking was kept to a minimum during the data entry task and was prohibited during completion of the fatigue test battery. Excluding lunches and breaks, actual daily work time was approximately 10.2 hrs. Actual work time included both continuous

TABLE 1

Activities during the Workday

0700–0800 hrs	DET* (Data Enry Task)
0800–0855	TBB (Test Battery and 10-Min Break)
0855–0955	DET*
0955–1050	TBB
1050–1115	DET
1115–1145	LUNCH
1145–1245	DET*
1245–1340	TBB
1340–1440	DET*
1440–1535	TBB
1535–1600	DET
1600–1630	LUNCH
1630–1730	DET*
1730–1825	TBB
1825–1900	DET

* Only scores from these one-hour sessions were used in the data analysis so that each error score was based on an equal number (1200) of entry trials.

data entry performance and the intermittent fatigue test battery.

On laboratory restdays the test battery was administered at the same times of day as on workdays. During these restdays subjects remained in the laboratory and engaged in relaxing activities (e.g., watching TV, playing board games, or reading) between test times.

Throughout the study subjects maintained their usual sleep times, eating habits, and caffeine use.

Data Analysis

A 5 (Workdays) × 5 (Test Times) × 2 (Subject Runs) analysis of variance (ANOVA) with repeated measures on workdays and test times was performed on the error scores of the data entry task from the second workweek of the study. This procedure tested for differences in error rate as a function of day of the week, time of day, and the interaction between day and time of day. The factor for subject run (a blocking factor) was added to account for variance attributable to arbitrary differences (e.g., seasonal variations, differences in experimenters, or within-group social interactions) between the two groups that potentially could affect performance. No hypotheses were tested against that factor.

A 7 (Two Restdays and Five Workdays) × 5 (Test Times) × 2 (Subject Runs) ANOVA with repeated measures on days and test times was performed on each score from most tasks of the fatigue test battery. This ANOVA design was used to test for (1) differences between workdays and restdays, (2) differences across the workweek, (3) changes across time of day, and (4) the interactions among these factors.

Since task demand was systematically varied in the grammatical reasoning and in the digit addition tasks, variations in task demand were tested by inserting an additional factor in the ANOVA. For the digit addition task a 2 (Task Demands—no constant versus constant) × 7 (Workdays) × 5 (Test Times) × 2 (Subject Runs) ANOVA was performed. For the grammatical reasoning task a 4 (Task Demands) × 7 (Workdays) × 5 (Test Times) × 2 (Subject Runs) ANOVA was performed. The four grammatical reasoning task demands were the three memory loads (2-, 4-, and 6-letter strings) and the four-letter reasoning task performed as part of the dual task.

Because choice reaction time was scored for each of four 2.5-min periods, a factor for periods was added to the ANOVAs for that task. Thus separate 4 (Periods) × 7 (Workdays) × 5 (Test Times) × 2 (Subject Runs) ANOVAs were performed on the reaction time, gaps, and error scores of that task.

RESULTS

Daily means for the second workweek (experimental week) and the two laboratory restdays are presented in Table 2 for all measures. F values testing for differences across days are also shown.

Work Simulation: Data Entry Task

Mean number of data entry errors per 60-min period for each day at each time of day are shown in Figure 1. As shown, there was a precipitous increase in errors at 1300 hrs on the fourth and fifth workdays. Errors also increased as a function of time of day. For example, there was a 19% increase in errors from 1500 to 1730 hrs. These results are supported by a statistically significant interaction of day by time of day, $F(16,108) = 2.06$, $p < 0.02$, and a significant main effect for time of day, $F(4,28) = 8.61$, $p < 0.001$.

Self-Report Measure: Stanford Sleepiness Scale

As shown in Table 2, subjective sleepiness and negative feelings significantly increased on workdays as compared with restdays. Positive feelings significantly decreased on workdays.

Performance Measures

Grammatical reasoning. Table 2 shows that overall response time was significantly faster during workdays and decreased from the first through the fourth workday. Response time also became faster as the day progressed, $F(4,28) = 4.01$, $p < 0.02$, as shown in Table 3. As expected, response time was slower for the 4- and 6-letter and dual-task versions as compared with the 2-letter version, $F(3,21) = 7.79$, $p < 0.002$, but there were no significant interactions among any of the factors.

Table 2 also shows that errors occurred more frequently on the first four workdays as compared with restdays and the final workday, $F(6,42) = 3.28$, $p < 0.01$. Errors also significantly increased across the day, $F(4,28) = 3.78$, $p < 0.02$, as shown in Table 2. There were no significant effects on error scores for task demand or any interactions.

Digit addition. As shown in Table 2, more correct additions were completed during the

TABLE 2

Daily Means across Days for All Dependent Measures

		Rest	Day 1	2	3	4	5	Rest	F	p
Data Entry errors	M	—	412	419	409	456	454	—	1.94	0.14
	SD	—	208	186	189	241	228	—		
Grammatical reasoning (mean of all versions)										
Response time (s)	M	3.04	2.80	2.78	2.43	2.29	2.50	2.62	5.96	0.001
	SD	0.86	1.08	1.15	1.04	1.10	1.13	1.03		
Errors	M	2.96	4.27	4.09	4.39	4.38	3.86	3.24	3.78	0.02
	SD	2.03	2.70	2.86	3.04	3.03	2.74	2.60		
Digit addition (mean of two versions)										
No. correct	M	64.12	60.93	62.28	61.81	60.98	65.49	66.56	4.54	0.002
	SD	14.42	15.16	15.91	16.81	17.62	14.96	14.27		
Errors (%)	M	4.91	7.73	7.96	8.36	10.70	7.47	5.92	2.06	0.08
	SD	4.20	8.98	6.69	7.81	12.34	8.22	5.37		
Time production										
Estimated/Target Time	M	0.90	0.89	0.91	0.88	0.89	0.88	0.84	0.38	0.89
	SD	0.19	0.22	0.24	0.21	0.23	0.18	0.19		
Reaction time										
Simple auditory										
RT (ms)	M	158	192	210	214	215	210	176	6.89	0.001
	SD	66	68	90	104	95	109	67		
Gaps (%)	M	1.19	2.19	2.89	4.07	2.22	2.77	1.52	1.21	0.32
	SD	2.58	3.99	4.33	9.28	2.95	6.47	3.37		
Dual auditory										
RT (ms)	M	269	277	329	319	311	294	269	4.68	0.001
	SD	105	84	135	144	118	98	89		
Gaps (%)	M	1.11	2.41	2.41	2.22	4.07	2.27	1.48	1.22	0.32
	SD	3.37	6.31	4.57	5.73	8.26	6.06	4.46		
Choice visual										
RT (ms)	M	585	629	615	628	606	600	602	1.53	0.17
	SD	89	89	121	144	126	120	87		
Gaps (%)	M	5.22	7.80	7.71	8.71	5.82	6.32	4.00	2.17	0.06
	SD	6.23	6.01	9.18	9.71	6.41	8.22	4.50		
Errors (%)	M	3.73	5.38	4.64	4.62	6.69	6.48	3.62	1.84	0.11
	SD	4.62	8.48	7.52	5.81	11.68	9.67	4.26		
2-Pt. Aud. discrimination (ms)	M	3.32	3.53	3.59	3.54	3.53	3.50	3.53	1.25	0.30
	SD	0.64	0.68	0.72	0.78	0.83	0.83	0.89		
Response alternation										
No. alternations	M	302	289	306	286	291	278	279	2.28	0.06
	SD	25	44	38	54	29	46	37		
Long alternations	M	77	87	70	76	79	80	84	0.69	0.66
	SD	38	37	40	47	50	46	47		
Hand steadiness										
Time off-target (%)	M	14.63	23.39	27.05	27.99	27.68	28.11	24.06	3.02	0.02
	SD	12.57	19.90	20.38	22.09	22.13	20.81	19.19		
Stanford sleepiness	M	1.98	2.91	3.02	2.96	3.13	3.26	2.40	5.12	0.001
	SD	0.78	1.14	1.08	0.88	1.14	0.98	0.81		
NPRU Adjective Checklist										
Positive	M	55.47	49.67	50.96	49.73	49.40	48.33	51.20	3.98	0.003
	SD	10.82	8.47	10.55	8.02	8.71	8.94	10.89		
Negative	M	14.29	17.22	18.16	18.22	18.84	18.84	16.02	4.69	0.001
	SD	4.12	4.74	5.65	4.16	5.00	4.39	4.23		
Oral temperature (C)	M	36.77	36.87	36.99	37.02	37.05	37.07	36.83		
	SD	1.49	1.42	1.69	1.39	1.53	1.73	1.85		

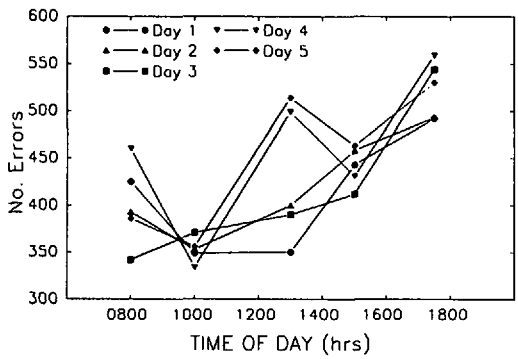

Figure 1. *Number of data entry errors per 60-min period as a function of time of day for each workday.*

restdays and the final workday. Significantly more additions were completed in the no-constant version of the task, $F(1,7) = 158.24$, $p < 0.001$, but effects for test time and the interactions were not significant. There were no significant effects for number of errors.

Time estimation. The ratio of produced interval to target interval significantly decreased across the day except for a rise at 1300 hrs, $F(4,28) = 3.15$, $p < 0.03$. No other effects were significant.

Simple auditory reaction time. As shown in Table 1, reaction time was significantly longer on workdays. No other effects were significant.

Gaps, or missed signals, were most frequent at 1000 hrs on the second and third days of the workweek. This result is supported by a significant Days × Test Times interaction, $F(24,168) = 1.88$, $p < 0.02$.

Dual-task performance auditory reaction time. It is apparent in Table 2 that reaction time was significantly longer on the second, third, and fourth workdays. No other effects were significant for this task.

TABLE 3

Grammatical Reasoning over Time of Day (Average of All Task Demands)

	Time (hrs)				
	0800	1000	1300	1500	1730
Response time (s)	2.71	2.70	2.73	2.61	2.46
Errors	3.76	3.62	3.58	4.03	4.44

Choice visual reaction time. Mean reaction times for the entire 10 min did not differ across days. However, reaction time increased with time on task, which was supported by a statistically significant effect for period, $F(3,21) = 7.23$, $p < 0.002$. There was also a significant Days × Period interaction in the reaction time score, $F(18,126) = 2.99$, $p < 0.003$. To illustrate this interaction, mean reaction times for the first and fourth periods are shown across workdays and restdays in Figure 2. As can be seen, first-period reaction times were 40–60 ms faster than fourth-period reaction times on restdays and the final workday. This difference was less than 15 ms on the other four workdays.

Choice errors increased significantly across the day, $F(4,32) = 3.65$, $p < 0.02$, and across periods, $F(3,24) = 5.29$, $p < 0.01$. There were no significant effects for gaps.

Two-point auditory discrimination. Fusion gap threshold significantly decreased as a function of time of day, $F(4,28) = 3.00$, $p < 0.04$. No other effects were significant.

Response alternation performance (tapping). Table 2 shows that the total number of alternations decreased in the latter half of the workweek and remained at that level on the final restday. This change, however, was not statistically significant, $p < 0.054$. No other effects were significant.

Hand steadiness. As shown in Table 2, time off-target was significantly elevated on workdays as compared with the first restday. It is also apparent that time off-target did not return to initial restday levels on the final restday. No other effects were significant.

Oral temperature. As shown in Table 2, oral temperature was significantly elevated on workdays. As expected, temperature also increased as a function of time of day to a peak

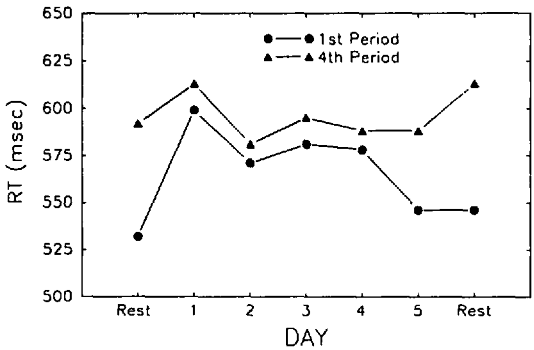

Figure 2. *First and last 2.5-min periods of choice reaction time over days.*

at 1500 hrs, before declining in the early evening. No other effects were significant.

DISCUSSION

Work-Rest Differences

Data entry work simulation. The results of the data entry task suggest that the work regimen was fatiguing. The number of data entry errors increased significantly as the 12-hr day progressed, being most frequent in the evening (1730 hrs). Errors also increased as the workweek progressed, being most frequent at 1300 hrs on the fourth and fifth workdays. The increase at this time can be interpreted as a fatigue-induced exacerbation of the "post-lunch dip" observed by other researchers (e.g., Colquhoun et al., 1968, 1969).

Self-report scales. The self-report scales of the test battery also indicated that the subjects felt the strain of the work routine. They reported feeling more tired and sleepy on the Stanford Sleepiness Scale on workdays as compared with restdays, and also reported less positive and more negative feelings on the NPRU Adjective Checklist on workdays.

Oral temperature. Work-rest differences were also apparent in body temperature, which was consistently higher on workdays as compared with restdays. Temperature also increased as the workweek progressed. We assume that the work activity produced this temperature increase, but we are not aware of any studies presenting similar results. The measure seems reliable, however, because (1) the increase in temperature across the day replicates previous findings (e.g., Aschoff et al., 1972) and (2) daily temperature means from the first practice/adaptation week were also higher on workdays. Unlike the second week, however, the first-week temperature tended to decrease as the week progressed. Reasons for the difference in trends across the two study weeks are not clear. We assume that the decrease across the first week was an adaptation effect and that the second week better reflects the work schedule.

Performance tasks. Fatigue indicated by progressively higher error scores at the job simulation and subjective reports of more weariness and strain during workdays was also reflected in many of the brief standardized performance tasks of the test battery. Fewer digit-addition problems were correctly completed; simple auditory reaction time was slower and more signals were missed (gaps); dual reaction time and choice reaction time (first 2.5-min period) were slower; and hand steadiness decreased. Response alternation (tapping) speed also appeared slower on workdays but did not reach statistical significance, $p < 0.06$. Time estimation and two-point auditory discrimination showed no work-rest differences.

Work-rest differences were also evident in the grammatical reasoning task, but the response time and error measures changed in opposite directions. Errors were more frequent on workdays, suggesting poorer performance, but response time decreased (became faster) on workdays, suggesting better performance. Consideration of the two measures together, however, indicates that the subjects sacrificed accuracy for greater speed on workdays. A separate analysis of response time in those sessions during which the subjects were most careful supports this hypothesis. Analysis of variance of reasoning sessions with four or fewer errors revealed no significant workday-restday differences in response time, $F(6,42) = 2.10$, $p < 0.08$. (According to the binomial test, the probability of nonrandom responding is less than 5% in those task sessions with fewer than five errors.)

The results of the grammatical reasoning task suggest that long workdays reduced motivation to perform carefully. In generalizing to the work situation, we speculate that a worker in a fatigued state could be more likely to take careless shortcuts to completion of a job. This potential sacrifice of safe work practices might be likely in tasks that are tedious because of high cognitive or information-processing demands, or those with extensive repetition (the reasoning task was performed four times at each test session).

Final workday improvements. For many of the tasks in the test battery, performance was poorest during the middle two or three days of the workweek. Performance on the last workday, when accumulated fatigue could be expected to be at its highest, often approached or achieved restday levels. On the fifth workday, grammatical reasoning errors decreased and response time increased (suggesting more careful responding). Also, digit-addition number correct increased, and both dual reaction time and choice reaction time (first 2.5 min only) decreased. This improvement in performance on a day when the worst performance might be expected is best explained by "end-spurt" or end-of-session effects. Given that the fifth workday was the final workday of the study and also the last day before the weekend, the anticipation of a reprieve from the work regimen could have motivated the subjects to perform better.

Final restday lack of improvement. Both motor tasks—hand steadiness and response alternation—failed to show either final workday or final restday improvements. In contrast, the other tasks and subjective scales in the battery returned to the prework restday levels. There is no clear explanation for this lack of improvement. The possibility that subjects were no longer motivated to perform well is not corroborated by their performance on the other tasks. It is possible that motor abilities had not recovered after two days of home rest following a 60-hr workweek. More than one restday, however, would be needed to track a reliable time course for recovery.

Diurnal Variations

Increases in data entry errors across the workday were reflected in two measures of the fatigue test battery—errors in the grammatical reasoning and choice reaction time tasks increased across the day.

Two tasks in the battery changed across the day in parallel with the increase in body temperature. Time production ratios (produced interval/real time) decreased across the day, indicating that the individual's "internal clock" was moving progressively faster with respect to real time. This result replicates previous findings (Pfaff, 1968). The perceived auditory discrimination gap also became smaller as the day progressed, suggesting more acute auditory sensitivity. Neither of these measures, however, showed work-rest differences.

Other tasks did not show significant time-of-day effects even though previous research (see Colquhoun, 1982) has demonstrated such effects with similar tasks. Most previous research, however, was designed to elucidate diurnal variations by controlling for fatigue and other intervening variables. The present study was designed to emphasize fatigue and may have eliminated time-of-day effects by opposing two antagonistic processes operating on the arousal system. From previous time-of-day research one would expect an improvement in performance across the day in parallel with the increase in body temperature (except, possibly, in the cognitive tasks). From a fatigue standpoint, however, one would expect a decline in performance with accumulated hours of work. The sum of these opposing processes could result in the observed lack of change—or highly variable change—in performance across the day.

Cognitive Demand

Previous research (Monk and Embrey, 1981) suggests that performance on tasks with high cognitive demand covaries negatively with the circadian rhythm of body temperature. We attempted to replicate these findings by systematically varying cognitive demand in the grammatical reasoning task. In that task cognitive demand was defined by the number of letters held in memory to be compared with the conditional statement. Performance was expected to peak earlier in the day at higher memory loads in opposition to the rise in body temperature. However, no interaction of task demand (memory load) with time of day was obtained, indicating no systematic difference in performance peak as a function of memory load. Overall response speed and error rate did increase with time of day, but, as noted earlier, this effect was best explained by the strategy of sacrificing accuracy for speed. That strategy is more likely a product of fatigue than of diurnal influences.

Taken together, these results imply that fatigue effects are more critical than circadian influences, at least within the day-shift hours observed in the present study. Circadian influences might be more critical on the night shift, where low arousal in the early morning would further detract from declining performance attributable to fatigue. Future laboratory and work-site studies will consider this issue.

In conclusion, several measures in the test battery were sensitive to work-rest differences in the present demonstration. The battery has potential, then, for detecting decrements in performance and alertness associated with hours of work at actual work sites. We are currently using the battery in two work site investigations of 12-hr days. Its flexibility in terms of selectivity and timing of tasks and its ease of administration are proving it to be an economical and effective way to collect round-the-clock data over periods of several weeks (see Rosa et al., 1986, for an introduction to these investigations).

REFERENCES

Aschoff, J. (1981). Circadian rhythms: Interference with and dependence on work-rest schedules. In L. C. Johnson, D. I. Tepas, W. P. Colquhoun, and M. J. Colligan (Eds.) *The twenty-four hour workday: Proceedings of a symposium on variations in work-sleep schedules* (DHHS [NIOSH] Publication No. 81-127, pp. 13–50). Washington, DC: U.S. Government Printing Office.

Aschoff, J., Geidke, H., Poppel, E., and Wever, R. (1972). The influence of sleep interruption and sleep deprivation on circadian rhythms in human performance. In W. P. Colquhoun (Ed.), *Aspects of human efficiency* (pp. 135–150). London: English Universities Press.

Baddeley, A. (1968). A 3-min reasoning task based on grammatical transformation. *Psychonomic Science, 10*, 341–342.

Bills, A. C. (1931). Blocking: A new principle in mental fatigue. *American Journal of Psychology, 43*, 230–245.

Colligan, M. J., and Tepas, D. I. (1986). The stress of hours of work. *American Industrial Hygiene Association Journal, 47*, 686–695.

Colquhoun, W. P. (1982). Biological rhythms and performance. In W. B. Webb (Ed.), *Biological rhythms, sleep, and performance* (pp. 59–86). New York: Wiley.

Colquhoun, W. P., Blake, M. J. F., and Edwards, R. S. (1968). Experimental studies of shift work II: Stabilized 8-hour shift systems. *Ergonomics, 11*, 527–546.

Colquhoun, W. P., Blake, M. J. F., and Edwards, R. S. (1969). Experimental studies of shift work III: Stabilized 12-hour shift systems. *Ergonomics, 12*, 856–882.

Dember, W. N., and Warm, J. S. (1979). *Psychology of perception*. New York: Holt, Rinehart & Winston.

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., and Dement, W. C. (1973). Quantification of sleepiness. A new approach. *Psychophysiology, 10*, 431–436.

Johnson, L. C., and Naitoh, P. (1974). *The operational consequences of sleep deprivation and sleep deficit* (AGARDograph No. 193, North Atlantic Treaty Organization). (Available from the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22151).

Johnson, L. C., Tepas, D. I., Colquhoun, W. P., and Colligan, M. J. (Eds.). (1981). *The twenty-four hour workday: Proceedings of a symposium on variations in work-sleep schedules* (DHHS [NIOSH] Publication No. 81–127). Washington, DC: U.S. Government Printing Office.

Monk, T. H., and Embrey, D. E. (1981). A field study of circadian rhythms in actual and interpolated task performance. In A. Reinberg, N. Vieux, and P. Andlauer (Eds.), *Night and shiftwork: Biological and social aspects* (pp. 473–480). Oxford, England: Pergamon.

Myers, J. L. (1979). *Fundamentals of experimental design.* Boston: Allyn and Bacon.

Naitoh, P. (1981). Circadian cycles and the restorative power of naps. In L. C. Johnson, D. I. Tepas, W. P. Colquhoun, and M. J. Colligan (Eds.), *The twenty-four hour workday: Proceedings of a symposium on variations in work-sleep schedules* (DHHS [NIOSH] Publication No. 81–127, pp. 693–720). Washington, DC: U.S. Government Printing Office.

Pfaff, D. (1968). Effects of temperature and time of day on time judgments. *Journal of Experimental Psychology, 76,* 419–422.

Rosa, R. R., and Colligan, M. J. (1986). *The NIOSH fatigue test battery* (NIOSH Internal Report). (Available from the National Institute for Occupational Safety and Health, Division of Biomedical and Behavioral Science, 4676 Columbia Pkwy., Cincinnati, OH 45226).

Rosa, R. R., Colligan, M. J., and Lewis, P. (1986). Extended workdays: Effects of 8-hour and 12-hour rotating shift schedules on test performance, subjective alertness, sleep patterns, and psychosocial variables. In *Proceedings of the Human Factors Society 30th Annual Meeting* (pp. 882–884). Santa Monica, CA: Human Factors Society.

Rosa, R. R., Wheeler, D. D., Warm, J. S., and Colligan, M. J. (1985). Extended workdays: Effects on performance and ratings of fatigue and alertness. *Behavior Research Methods, Instruments, and Computers, 17,* 6–15.

Rutenfranz, J., Colquhoun, W. P., Knauth, P., and Ghata, J. N. (1977). Biomedical and psychosocial aspects of shift work. *Scandinavian Journal of Work Environment and Health, 3,* 165–182.

Volle, M., Brisson, G. R., Perusse, M., Tanaka, M., and Doyon, Y. (1979). Compressed work-week: Psychophysiological and physiological repercussions. *Ergonomics, 22,* 1001–1010.

Wever, R. A. (1981). On varying work-sleep schedules: The biological rhythm perspective. In L. C. Johnson, D. I. Tepas, W. P. Colquhoun, and M. J. Colligan (Eds.), *The twenty-four hour workday: Proceedings of a symposium on variations in work-sleep schedules* (DHHS [NIOSH] Publication No. 81–127, pp. 51–86). Washington, DC: U.S. Government Printing Office.

Williams, H. L., and Lubin, A. (1967). Speeded addition and sleep loss. *Journal of Experimental Psychology, 73,* 313–331.