ELSEVIER International Journal of Industrial Ergonomics 21 (1998) 307–321

International Journal of Industrial Ergonomics

Some effects of 8- vs. 10-hour work schedules on the test performance/alertness of air traffic control specialists

David J. Schroeder[a,*], Roger R. Rosa[b], L. Alan Witt[c]

[a] *FAA Civil Aeromedical Institute, Human Resources Research Division, Oklahoma City, OK, USA*
[b] *National Institute for Occupational Safety and Health, Division of Biomedical and Behavioral Science, Cincinnati, OH, USA*
[c] *Barnett Banks, Jacksonville, Florida, USA*

Received 27 October 1995; accepted 10 January 1997

Abstract

A 10 h 4 d rotating shift schedule worked by some Air Traffic Control Specialists (ATCSs) was compared to the more traditional 8 h 2-2-1 rapidly rotating schedule. Measures of performance and alertness were obtained from a group of 52 ATCSs at an en route ATC center on tasks in the NIOSH fatigue test battery. Additional information on sleep patterns, mood, and somatic complaints was also gathered. Results confirm that tests comprising the NIOSH battery are sensitive to fatigue and diurnal variations associated with a rotating shift schedule. Test performance of ATCSs on the 10 h shift did not differ from those on the 8 h schedule for any of the parameters, when comparing the initial 4 d of the work week. Test performance was notably poorer on the night shift that occurred on the final (fifth) day of the 2-2-1 8 h schedule. For both schedules, there was evidence of changes in alertness on some of the NIOSH performance measures within work days and across days of the week. Changes in test performance and mood ratings corresponded to the decline in self-reported sleep time across the work week.

Relevance to industry

The increased emphasis on compressed work weeks within industry and other settings has raised a number of issues concerning how longer work days impact sleep, fatigue, mood, and performance. There have been relatively few attempts to systematically utilize field studies to gather relevant data so that managers have a scientific basis for decision-making. This study is one example of an approach to developing a more effective data base for decision-making regarding both a rapidly rotating shift schedule and a compressed work schedule.

Keywords: Shift work; Fatigue; Human performance; Air traffic control; Mood; Sleep; Computerized tests

1. Introduction

During the past 10 yr, management has been faced with increased employee demands for more flexible work schedules, including interest in

* Corresponding author.

PII S0169-8141(97)00044-9

"compressed" work schedules. A "compressed" work schedule refers to any work week where employees are allowed to complete their work in four or fewer days. Numerous questions have been raised concerning the possible effects of compressed schedules on productivity, job efficiency and fatigue, and associated concerns with safety and health. Duchon and Smith (1993) provide a comprehensive overview of many of the performance issues associated with extended workdays.

1.1. Impact of compressed work schedules

Empirical research on the impact of compressed work schedules has focused more closely on employees' subjective reports concerning fatigue, alertness, mood, job satisfaction and conflicts with family activities, and leisure time. Outcomes have indicated: (a) increases in organizational effectiveness (Hartman and Weaver, 1977; Wheeler, 1970) as well as no increases (Calvasina and Boxx, 1975); (b) increased satisfaction brought about by more leisure time (Hodge and Tellier, 1975; Steele and Poor, 1970) but greater fatigue, conflict with evening activities, and conflict between the work schedule and family and child-related activities (Hodge and Tellier, 1975; Kenny, 1974); (c) a full range of positive, negative, and neutral affective responses (cf. Dunham et al., 1987); and (d) both greater fatigue (e.g., Goodale and Aagaard, 1975; Hodge and Tellier, 1975) and no differences in fatigue (Latack and Foster, 1985).

Changes in performance and alertness associated with compressed work schedules have, until recently, received less attention. Of the performance-based studies, nearly all involve comparisons of 8 h and 12 h shift schedules. In an early exception, Volle et al. (1979), reported that factory employees on the 10 h vs. 8 h schedule did not differ significantly on reaction time but did display decreased grip strength and higher critical flicker fusion (CFF) thresholds. However, the authors concluded that the increase in fatigue remained within acceptable limits and that there was no evidence that these changes affected overall productivity in the manufacturing plant. Peacock et al. (1983) on the other hand, found in a study of police officers, improved subjective alertness, sleep, and cardiovascular fitness (12 h vs. 8 h). No significant differences were noted on CFF thresholds or grammatical reasoning tests. Mills et al. (1983), noted that employees on a 12 h shift schedule evidenced significant increases in subjective fatigue and grammatical reasoning errors from start to completion of the work day. However, a majority of the increase in errors occurred between the 1st and 6th hours of the 12 h work day. Nurses performed more rapidly on the grammatical reasoning test across the work day and expressed high levels of satisfaction with the 12 h schedule. Daniel and Potasova (1989) also reported some differences between 12 h and 8 h personnel on several cognitive and psychomotor tasks; however, these findings may have been influenced by differences in initial performance capabilities between the 2 groups. Lewis and Swaim (1988), using a number of measures of employee performance and fatigue, compared the effects of 8 h and 12 h shift schedules at an experimental nuclear reactor. While the results were mixed, with some indications of greater fatigue on the 12 h schedule, direct on-the-job performance measures favored the 12 h schedule. A vast majority of the employees favored the 12 h schedule and the authors concluded that the 12 h shift schedule was a "reasonable alternative to an 8 h schedule (Lewis and Swaim, p. 513)."

1.2. NIOSH fatigue test battery

The computerized National Institute of Occupational Safety and Health (NIOSH) Fatigue Test Battery was developed to quantify changes in several indices of cognitive, sensory, and perceptual-motor performance and self-reported subjective feelings associated with shift work. As part of that development, Rosa et al. (1985) assessed differences in the test performance of subjects working 6 8 h days or 4 12 h days. They found that individuals on the 12 h 4 d work week reported greater fatigue on several of the self-report measures than when on a 8 h 6 d work week. Greater evidence of fatigue was also found on the grammatical reasoning and digit addition cognitive performance measures from the test battery. In a subsequent laboratory investigation to assess the effects of fatigue and diurnal variations on performance on the test battery, Rosa and Colligan (1988) compared the effects of working 5 12 h days in the laboratory, using a data entry job simulation task with rest periods. They found that changes in performance on the data entry task associated with the work day and work week corresponded closely with subjective ratings and performance on a number of the tasks comprising the test battery. Rosa and Colligan (1988) concluded that the NIOSH Fatigue Test Battery is sensitive to long hours of work and to the influence of circadian rhythms on performance. In field studies, Rosa et al. (1989) and Rosa and Bonnet (1993) found evidence of significant differences in self-reported sleep time and fatigue, as well as performance on some aspects of the test battery, when comparing employees who were working 8 h and 12 h shift schedules at gas utilities and continuous processing plants. Thus, the findings of Rosa and his colleagues confirm that employees working a 12 h compressed work schedule experience greater fatigue and exhibit lower performance capabilities on some test measures as compared to those on more traditional 8 h schedules. In a 3 to 5 year follow-up, Rosa (1991) found that the sleep loss and performance declines were still present in employees on the 12 h shift schedule. However, employees continued to express generally high levels of satisfaction concerning the 12 h shifts and there was no operational evidence that safety was compromised by the associated fatigue. The lack of any demonstrable change in the operational performance measures may be due to the fact that the performance measures are not sufficiently rigorous to detect the effects of fatigue or that the performance requirements in the operational environment do not require as quick performance as is measured in the various NIOSH fatigue tests.

1.3. FAA work schedules

The FAA has approved the use of compressed and flexible rotating work schedules for its employees, including air traffic control specialists (ATCSs). This action included temporary approval for the use of 10 h work days. While FAA management closely reviewed various ATCS performance parameters to identify possible negative effects from the 10 h schedule, they also decided that a scientific study should be undertaken to assess the potential effects of working the 4 d 10 h shift schedule on employee performance capabilities. Since there is little information available in the literature concerning the 10 h work day, and the ATC work environment is sufficiently unique from the work environments included in the above-mentioned studies, this study was initiated to compare the effects of the existing 2-2-1 8 h rotating shift schedule with that of the 4 d 10 h rotating schedule on measures of employee cognitive performance and self-reported sleep and mood.

2. Method

2.1. Measurements

2.1.1. NIOSH fatigue test battery

This flexible, computerized test battery was developed by Rosa et al. (1985) and Rosa and Colligan (1988) specifically for applications in field experimentation with employees working on different shift schedules. Users can select from a group of tests that assess cognitive, perceptual-motor, and motor skills. Additional tests and self-report measures of alertness, fatigue, and the quality and duration of sleep can be incorporated into the battery, with limited programming requirements. Flexibility is also provided by the ability to tailor the test length to the research requirements and available time. The investigator is thus able to construct a test battery that is highly responsive to the job demands and requirements of the work setting. The choice reaction time, mental arithmetic, and grammatical reasoning tests were selected for inclusion in this study both on the basis of their demonstrated sensitivity to alterations in alertness and association with the job tasks of an ATCS. The relevance of these tasks to the ATC occupation is further supported by recent findings of Broach and Aul (1993), who used interviews of ATCSs and subsequent ratings on the Position Analysis Questionnaire to identify attributes of abilities or aptitudes required of ATCSs. Of greater relevance were perceptual speed, closure, reaction time, and short-term memory. Numerical computation, arithmetic reasoning, and convergent and divergent thinking were also somewhat more relevant for the ATC profession than for other jobs.

The choice reaction time task consisted of random presentation of the words TRUE or FALSE on a VDT for a total of 150 trials over approximately 10 m. The intertrial interval was random, with a range of 2–5 s. Subjects were required to press a push-button switch labeled "TRUE" or "FALSE" on a specially developed response box as quickly as possible to indicate the correct word. The ATCSs forefinger and middle finger of his/her preferred hand rested on the buttons during the trial. For this study, scores for the CRT task included the mean reaction time and number of errors (i.e. incorrect responses).

The mental arithmetic test is an adaptation of the test developed by Williams and Lubin (1966). At the beginning of the task, a randomly selected constant between the values of 3 through 9 was presented for 3 s and then removed for the remainder of the task. ATCSs were required to add the constant to the sum of 2 single digits and then type the last digit of the overall sum on the keyboard. The digits varied across trials and were generated immediately after an ATCS's response. Scores for the task included the number correct and number of errors during the 3 m time period.

The grammatical reasoning task is a variation of the well-known task first devised by Baddeley (1968). In this 16-trial task, a 3-letter stimulus string (e.g., JLN) was presented for 2 s, removed, and then followed after 3 s with a conditional statement such as "J DOES NOT PRECEDE N". The ATCS was required to press a push-button switch labeled "TRUE" or "FALSE" as quickly as possible to indicate whether the statement described the letter string. Scores for the GR task included average response latency for correct responses and total number of errors.

At the beginning of each testing session, subjects responded to 10 choice reaction time trials and 60 s of digit addition. These mini-sessions served the dual purpose of providing a "warm-up" and resolving any potential software or hardware problems before commencing the full battery.

2.1.2. Daily sleep, somatic complaints, and mood

The test battery was programmed to include items about amount of sleep, ratings of quality of sleep, mood, and somatic complaints. Subjects were asked to indicate their time of retiring, arising, sleep latency, and number of awakenings. A series of 5 questions, utilizing 5-point Likert rating scales, were included to evaluate the depth and quality of sleep. Workers also responded to the 29 item (19 positive and 10 negative) Naval Psychiatric Research Unit (NPRU) mood scale (Johnson and Naitoh, 1974). ATCSs also provided ratings of workload, using the Task Load Index (TLX) scale developed by National Aeronautics and Space Administration (NASA). Results of the workload data are not included in these analyses. Respondents were asked to indicate the presence or absence of each of 19 potential somatic complaints (e.g. headache, back pain, etc.).

2.2. Subjects

A total of 56 ATCSs from an en route air traffic control center initially volunteered to participate in the study. Subjects were predominately male (86%), ranging in age from 28 to 50 (mean age = 37.9). Their experience as ATCSs ranged from 4 to 22 yr. Since the opportunity to work the 10 h schedule at this facility was based on seniority, ATCSs on the 10 h schedule were approximately 4 yr older than those on the 8 h schedule. Prior to the initiation of the study, each ATCS was provided a description of the proposed study and asked to sign a consent form concerning the research project. A numerical code was assigned to each subject for the test sessions to ensure anonymity. Of the group of 56, 26 ATCSs working the 8 h 2-2-1 schedule and 26 on the 10 h 4 d schedule completed a sufficient number of sessions (10 or more) to be included in the study.

2.3. Rotating work schedules

The 8 h and 10 h rotating work schedules are illustrated in Fig. 1. Under the 8 h rapidly rotating, phase advancing schedule, ATCSs worked 2 consecutive afternoons, 2 mornings, and then returned on a quick turnaround (less than 10 h) to work a night shift. This schedule has been in use in ATC facilities for many years, and a considerable body of research in the 1970s was dedicated to evaluating the 2-2-1 8 h schedule vs a straight 5 d 8 h rotating

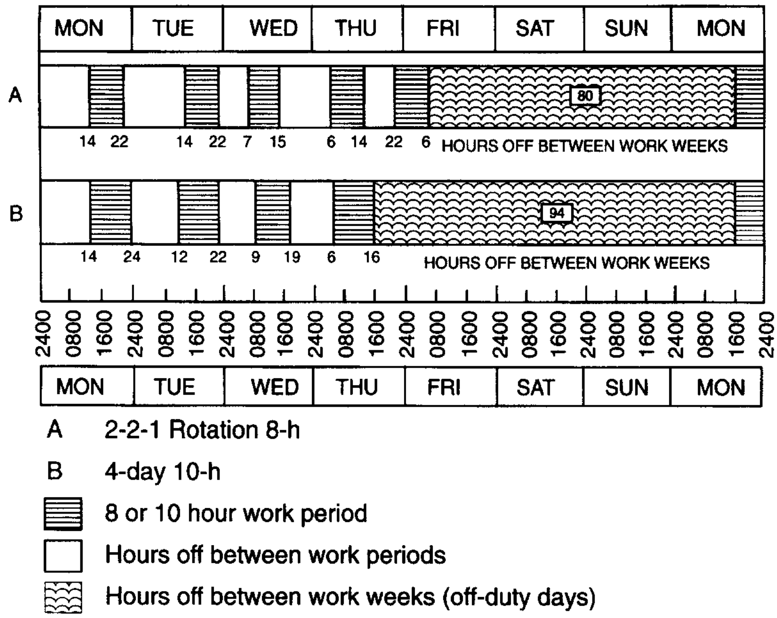

Fig. 1. A graphic representation of a work week on the 2-2-1 8 h and 4 d 10 h shift schedules. The weeks shown begin on Monday; however, an ATCS's work week may start on any day and the week end will come on days other than Saturday and Sunday.

schedule (Melton et al., 1971, 1973, 1975; Saldivar et al., 1977). On the 2-2-1 schedule, there are 2 nights when the time between the end of one shift and the beginning of another is sufficiently short to reduce the amount of available sleep time. The average number of hours between the end of the work day on day 1 and the start of the work day on day 2, and between day 3 and 4 was approximately 15; the average was approximately 9.3 h between day 2 and 3, and 8.1 h between day 4 and 5. ATCSs on the 10 h schedule also had variable starting times across the 4 d; working 2 afternoons followed by 2 mornings. On the 10 h schedule, the average number of hours between the end of a workday and the start of the next was: 12.2 between day 1 and 2, 11.1 between day 2 and 3; and 11.2 between day 3 and 4.

2.4. Test schedule

The NIOSH Fatigue Test Battery was administered on 3 occasions during the course of each work day over a three week time period. The initial session was conducted at the time the ATCS arrived at the facility. Session 2 was completed 2 h prior to the end of the work day (at the end of 6 h for the 8 h participants and 8 h for the 10 h participants). The third and final session was administered at the close of the work day. To the extent possible, the test sessions for each ATCS were administered at the same time across each of the 3 weeks. Some disruption in the test schedule occurred for one of the groups of 2-2-1 8 h and 10 h subjects as a result of a snow storm, which restricted travel to and from the ATC en route facility for several days.

2.5. Procedure

Seven microcomputers for administering the NIOSH fatigue test battery were located in a separate room within the en route center. The computers contained all of the instructions for completing each test session, along with the performance tests and rating scales. All response data were also stored on a computer. Following introduction to the computers and the test battery, test sessions were self-administered. An experimenter was available throughout the testing period to respond to questions and to intervene if a problem occurred with the computer.

2.6. *Test battery data analysis*

2.6.1. *Data exclusions*

Data from overtime days (only 5 or 6 d total), and dubious performance sessions were excluded from statistical analyses. Choice reaction time scores were excluded if more than 50 errors occurred. Digit addition scores were excluded if more than 45 errors occurred. Grammatical reasoning scores were excluded if more than 7 errors (i.e., 50% or more errors) occurred. These exclusion criteria are consistent with procedures used in previous studies associated with the NIOSH Fatigue Test Battery.

There were only 12 problematic scores for the choice reaction time measure. For digit addition errors, from 0–14% of the responses were excluded. There was no evidence that the problematic responses for those measures were systematically related to time of day or day of the week. A slightly higher proportion of the grammatical reasoning scores were excluded on the basis of a high number of errors (0–18%) or unusually fast response times (4–21%). For the latter measure, there was a significant interaction between day and session, based on an ANOVA calculated only on scores from the 8 h group ($F(2,192) = 2.78$, $p < 0.01$). With this exception, there was no evidence of any differences in problem scores between subjects from the two work schedules. The higher percentage of problematic scores on the grammatical reasoning test is consistent with other outcomes. In general, the ATCSs who participated in this study performed more accurately on the computer-based tasks than did other work groups (clerical/office personnel, control room operators, and gas control workers, Rosa and Colligan (1988); Rosa and Bonnet (1993) and Rosa et al. (1989)).

2.6.2. *Data transformations*

Several of the dependent variables in the test battery were transformed to approximate a normal distribution. Grammatical reasoning response time and choice reaction time were transformed to their inverses. Grammatical reasoning errors, choice reaction time errors, and digit addition errors were analyzed as percentage scores transformed to the arcsine of their square roots. The procedures for transforming the data were consistent with those used in previous studies and are based on the recommendations of Myers (1979).

2.6.3. *Analysis of variance*

Prior to analyses, data were collapsed by work day and test session over the three weeks. The effects of shift schedule, workdays, test sessions within workdays, and their interactions were tested for statistical significance with analysis of variance (ANOVA). For these *between* schedule comparisons, day 5 of the 8 h shift schedule (night shift) was excluded from the ANOVAs. Because of unequal cell frequencies, least-squares regression solutions to the ANOVAs were computed using the SAS General Linear Models Procedure. In addition, supplementary repeated-measures ANOVAs were performed *within* each schedule (i.e., excluding any data from the other shift schedule) testing the effects of workdays, sessions, and their interaction. These supplementary ANOVAs eliminated between-group variance to obtain a more powerful test of changes within a shift schedule. Day 5 was included in the within-schedule ANOVAs for the 8 h schedule. Post hoc tests of differences between means for those variables which were significantly different across days, sessions, and the interactions of days and sessions were determined with the Newman–Keuls test. An alpha level of $p < 0.05$ was considered statistically significant for all analyses.

3. Results

3.1. *NIOSH fatigue test battery*

Average scores on the three components of the NIOSH test battery across days of the work week are presented in Table 1. Results of the ANOVAs and post-hoc comparisons are also presented in Table 1. Of the various comparisons, there were no instances where differences in NIOSH test performance between ATCSs on the 8 and 10 h schedules were statistically significant. Performance related differences were generally due to effects associated

Table 1
Average performance scores for each component of the NIOSH test battery across days of the work week

	Day of the work week					ANOVA	Post-hoc results[a]
	1	2	3	4	5		
Choice reaction time							
RT (s)							
8 h	0.432	0.441	0.466	0.467	0.482	$F(4,106) = 7.59$, $p < 0.0001$	1,2 < 3,4,5;3 < 5
10 h	0.417	0.421	0.442	0.439		$F(3,77) = 2.96$, $p < 0.04$	1,2 < 3,4
Average	0.425	0.431	0.454	0.453		$F(3,156) = 8.13$, $p < 0.001$	1,2 < 3,4
No. errors							
8 h	3.58	3.56	2.97	3.43	4.09	NS	NS
10 h	2.49	2.49	2.42	2.62		NS	NS
Average	3.05	3.03	2.69	3.02		NS	NS
Digit addition							
No. attempted							
8 h	97.70	100.43	101.65	95.96	93.69	NS	NS
10 h	100.71	102.58	103.07	101.68		NS	NS
Average	99.19	101.59	102.39	98.89		$F(3,155) = 3.81$, $p < 0.02$	1 < 2,3 > 4
No. errors							
8 h	3.02	2.59	3.21	3.34	4.27	NS	NS
10 h	3.12	3.13	3.04	3.12		NS	NS
Average	3.07	2.87	3.12	3.23		NS	NS
Grammatical reasoning							
Response time (s)							
8 h	2.97	2.90	2.79	2.95	2.92	NS	NS
10 h	2.87	2.78	2.77	2.70		NS	NS
Average	2.91	2.84	2.78	2.82		$F(3,149) = 2.89$, $p < 0.04$	1 > 3
No. errors							
8 h	2.01	1.91	1.85	1.71	2.22	NS	NS
10 h	1.93	1.52	1.54	1.64		NS	NS
Average	1.97	1.71	1.69	1.68		NS	NS

[a]Significant mean differences were determined with the Newman–Keuls test.

with day of the work week, sessions, and the day-by-sessions interactions.

3.1.1. Choice reaction time

ATCSs on both the 8 and 10 h schedules exhibited a steady increase in choice reaction times from the initial to final work day during their respective work weeks, with the 10 h group exhibiting slightly quicker overall average reaction times. For both work schedules, the average reaction times on the first two days of the work week were significantly quicker than on days 3 and 4. For the 8 h schedule, the average reaction time on the 5th day of the work week was significantly slower than on any of the previous 4 d. Since days 3 and 4 were always morning shifts and the 5th day of the 8 h schedule was always a night shift, we were unable to statistically separate out the extent to which the changes in reaction times were attributable to fatigue arising from the day of the work week or time of day the tests were administered. The slowest reaction time occurred for ATCSs during the night shift, where the mean reaction time (0.482 s) was approximately 12% above the average noted on the first day of the work week.

Session-related changes in the performance measures are presented in Table 2. While the effect of session on choice reaction time was significant

Table 2
Average performance scores for each component of the NIOSH test battery across sessions

	Session[a]				
	1	2	3	ANOVA	Post-hoc results[b]
Choice reaction time					
RT (s)					
8 h	0.453	0.463	0.455	$F(2,54) = 4.24$, $p < 0.02$	NS
10 h	0.433	0.436	0.420	$F(2,54) = 10.93$, $p < 0.001$	1,2 > 3
Average	0.442	0.446	0.433	$F(2,108) = 13.81$, $p < 0.001$	1,2 > 3
No. errors					
8 h	3.22	3.21	4.19	$F(2,54) = 11.14$, $p < 0.001$	1,2 < 3
10 h	2.23	2.46	2.85	$F(2,54) = 4.63$, $p < 0.02$	1 < 3
Average	2.74	2.86	3.26	$F(2,108) = 9.28$, $p < 0.001$	1,2 < 3
Digit addition					
No. completed					
8 h	97.35	97.85	98.46	NS	NS
10 h	101.69	102.13	102.33	NS	NS
Average	99.76	100.40	101.43	NS	NS
No. errors					
8 h	3.14	3.12	3.63	NS	NS
10 h	3.06	2.91	3.34	NS	NS
Average	2.96	2.93	3.34	NS	NS
Grammatical reasoning					
Response time (s)					
8 h	2.97	2.91	2.83	$F(2,52) = 8.40$, $p < 0.001$	1 > 3
10 h	2.87	2.77	2.70	$F(2,54) = 14.62$, $p < 0.001$	1 > 2,3
Average	2.92	2.83	2.76	$F(2,105) = 24.48$, $p < 0.001$	1 > 2 > 3
No. errors					
8 h	1.94	1.96	1.93	NS	NS
10 h	1.57	1.77	1.63	NS	NS
Average	1.70	1.83	1.77	NS	NS

[a]Means for the 8 and 10 h sessions respectively includes a 5 and 4 d work week. Values for the averages are based on a 4 d work week.
[b]Significant mean differences were determined with the Newman–Keuls test.

for both the 8 and 10 h schedules, post hoc comparisons were significant only for the overall and 10 h comparisons. The average choice reaction times for the first two sessions were significantly slower than the final session of the work day.

The significant day by session interaction ($F(8,211) = 3.17$, $p < 0.02$) for choice reaction time by ATCSs on the 8 h schedule is illustrated in Fig. 2. Through the first three days of the work week, average reaction times were faster on the final session. However, on day 5 (night shift) the average reaction time increased from 0.460 at the start of the work day to 0.495 during the final session. Post hoc tests among means indicated that these differences were significant.

For the 8 h schedule, the average number of errors for sessions 1 and 2 were fewer than the final session. While ATCSs in the 10 h group evidenced a similar increase in errors from the first through the third session, post hoc comparisons revealed that only the difference between the first and third session was statistically significant. Fig. 2 illustrates the significant day by session interaction in the choice reaction time error measure for the 8 h group ($F(8,211) = 3.70$, $p = 0.001$). During the first two days of the work week, average reaction time

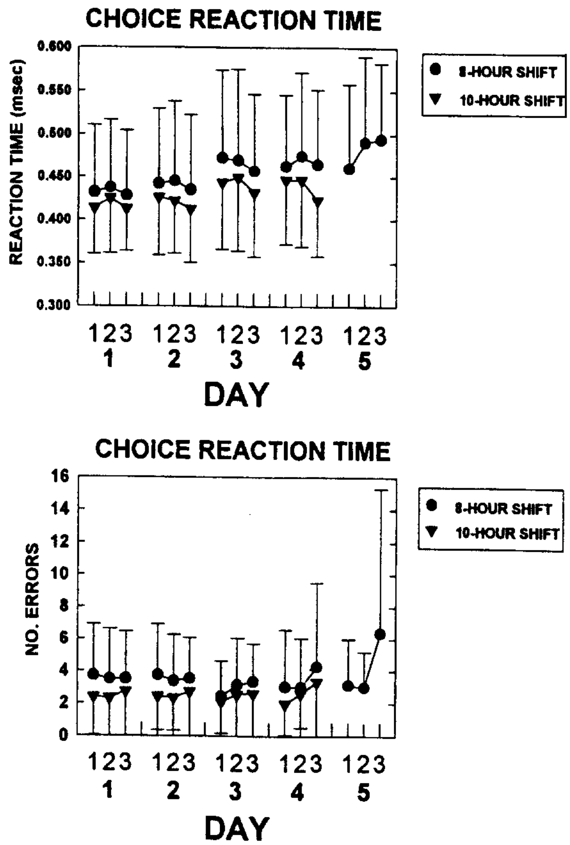

Fig. 2. Means and standard deviations for the choice reaction times and average number of reaction time errors for each session of the day across the work week, for ATCSs working the 2 shift schedules.

errors for the second and third session were below those of the first session. On subsequent days, average choice reaction time errors for the final session were consistently above that of the first session. The difference became more pronounced as the week progressed.

3.1.2. *Digit addition*

Effects of the 2 shift schedules on ATCS performance on the digit addition task (number attempted and number of errors) are presented in Tables 1 and 2. Across days of the work week, there was a general increase in number of pro-blems attempted. While the *between* schedule comparison revealed a significant day effect for the number of errors measure, none of the subsequent *within* schedule comparisons reached statistical significance. Post hoc comparisons across the first four work days revealed that the number of problems completed on days 2 and 3 was significantly greater than on day 1. Fewer problems were completed on day 4 than either day 2 or 3. The average number of errors remained relatively stable across the first 4 d of the work week for both the 8 and 10 h groups, the average for the 8 h group on day 5 was above that of any of the preceding days.

3.1.3. *Grammatical reasoning*

Average performance of ATCSs on the grammatical reasoning test (response time and errors) across days of the work week and sessions are presented in Tables 1 and 2. The significant *between* schedule day effect is associated with the general decline in response time for ATCSs in both groups from the first day through day 3 of the work week. Post-hoc results indicate that the average for day 3 was significantly slower than on day 1. On the fourth day, ATCSs on the 10 h schedule exhibited response times that were quicker than those on day 3, while ATCSs on the 8 h shift had slightly slower response times than on day 3. On the fifth day (night shift), the average response times of ATCSs on the 8 h shift were comparable to those noted on the second and fourth days. Overall, ATCSs had quicker response times during the final session than on the first two sessions of the work day.

3.2. *Daily mood and sleep*

3.2.1. *Mood ratings*

Average levels of positive and negative moods associated with work are presented in Table 3. The *between* shift ANOVA for positive mood ratings yielded significant effects for the day and day by session interaction $F(6,310) = 10.02$, $p = 0.001$. Positive ratings of mood remained relatively stable across the first 2 d of the work week for both groups. Ratings on days 3 and 4 were significantly below those provided on days 1 and 2. Post hoc comparisons for days of the work week for ATCSs on the 10 h schedule did not reach statistical significance. Ratings for ATCSs in the 8 h group declined

Table 3
Average ratings of mood and quality of sleep and amount of sleep across days of the work week and sessions

	Day of the work week						
	1	2	3	4	5	ANOVA	Post-hoc results[a]
NPRU mood scale							
Positive rating							
8 h	50.97	51.54	49.89	47.32	41.49	$F(4,106) = 22.48$, $p < 0.001$	2,3 > 4 > 5;1 > 4 > 5
10 h	52.37	52.02	50.40	50.99		NS	NS
Average	51.65	51.78	50.15	49.14		$F(3,156) = 4.33$, $p < 0.006$	1,2 > 3,4
Negative rating							
8 h	15.57	15.14	15.88	15.85	18.63	$F(4,106) = 13.45$, $p < 0.001$	1-4 < 5
10 h	13.83	13.86	14.97	14.67		$F(3,77) = 2.93$, $p < 0.04$	NS
Average	14.73	14.51	15.41	15.26		$F(3,156) = 3.19$, $p < 0.03$	2 < 3
Total sleep							
8 h	8.06	7.89	6.03	5.82	3.57	$F(4,105) = 67.60$, $p < 0.001$	1,2 > 3,4 > 5
10 h	8.06	7.15	6.97	5.72		$F(3,75) = 18.21$, $p < 0.001$	1 > 2,3 > 4
Average	8.06	7.53	6.52	5.77		$F(3,154) = 39.77$, $p < 0.001$	1 > 2 > 3 > 4
Awaken refreshed							
8 h	3.28	3.23	2.82	2.75	2.41	$F(4,105) = 6.41$, $p < 0.001$	1 > 3,4,5;2 > 4,5
10 h	3.46	3.40	2.97	2.93		$F(3,76) = 6.80$, $p < 0.001$	1,2 > 3,4
Average	3.37	3.31	2.90	2.83		$F(3,155) = 9.38$, $p = 0.001$	1,2 > 3,4

	Session				
	1	2	3	ANOVA	Post-hoc results[a]
NPRU mood scale					
Positive rating					
8 h	49.11	48.90	46.63	$F(2,54)$ 6.38, $p < 0.004$	1,2 > 3
10 h	51.87	51.54	50.86	NS	NS
Average	50.94	51.14	49.94	NS	NS
Negative rating					
8 h	15.66	16.15	16.88	$F(2,54)$ 6.60, $p < 0.003$	1 > 3
10 h	14.15	14.20	14.67	NS	NS
Average	14.77	14.85	15.31	NS	NS

[a]Significant mean differences were determined with the Newman–Keuls test.

for both days 4 and 5, with both of the ratings significantly different from the ratings made on any of the succeeding days.

Session related changes in the positive and negative moods also appear in Table 3. On the afternoon shifts (days 1 and 2), positive mood ratings declined from start to close of the work day. Post hoc comparisons indicate that for the first afternoon the first two ratings were significantly higher than for the final rating of the day. For the second day, the first rating was significantly higher than either the second or final rating. For the morning shifts, ratings for the final 2 sessions of the work day were above those of the start of the day. While the differences were not significant on the first morning, on the second, the first rating was significantly

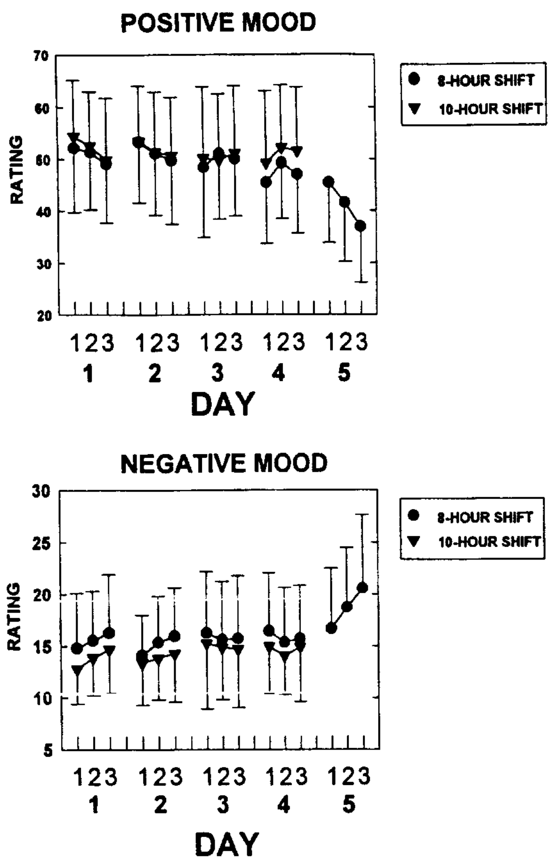

Fig. 3. Means and standard deviations for positive and negative mood ratings for each session of the day across the work week, for ATCSs working the 2 shift schedules.

lower than the second rating. Positive ratings for the night shift evidenced a significant decline from start to close of the work day (each successive rating was less positive.

As is evident in Fig. 3, changes in ratings of negative mood tend to mirror those noted for positive mood. Negative mood remained relatively stable across the first 4 d of the work week, only the ratings of day 2 and day 3 differed significantly from each other. The highest negative mood rating (18.63) occurred during the night shift (day 5), which was significantly higher than the average on any of the preceding 4 d. The largest change in mood across sessions occurred for ATCSs involved in the night shift, where positive mood declined and negative mood increased.

3.2.2. Sleep ratings

Means for the total sleep time and ratings of feeling refreshed following sleeping are presented in Table 3. The ANOVA for the sleep diary data yielded significant effects for the day and shift by day interaction. Both the 8 h and 10 h ATCSs exhibited a general decline in total sleep time from an average of 8.06 h on the evening prior to the first day of the work week to approximately 5.77 h on the evening prior to the fourth day of the work week. Overall post-hoc comparisons indicated that the average amount of sleep was significantly less than the previous day for each day of the work week. ATCSs on the 8 h shift exhibited the lowest average number of hours of sleep on the day prior to the night shift (3.57 h).

Changes in subjective ratings of feeling refreshed following sleep corresponded to the changes noted in the amount of sleep. The ANOVA revealed a significant effect for day. Feeling refreshed declined from an average rating (based on a scale of 1 to 5) of 3.28 (8 h) or 3.46 (10 h) for the evening prior to the first day of the work week to 2.75 and 2.93 for the evening prior to the fourth day. Across the two schedules, ratings for day 1 and 2 were significantly higher than on days 3 and 4. The lowest rating was that of 2.41 for sleep that occurred during the day for ATCSs on the 8 h schedule prior to the night shift. While that value is not significantly different from the rating provided on day 4, the ratings for day 4 and 5 were significantly below the ratings provided on days 1 and 2. Even though ratings for most of the other quality of sleep questions evidence a general decline from the first through final day of the work week, the overall differences were less prominent than those for "feeling refreshed following sleep".

3.2.3. Somatic complaints

ATCSs involved in this study reported very few somatic complaints, an average of less than 2 complaints per individual per session. There was no evidence of any significant changes in the somatic complaints across either days of the work week or sessions.

4. Discussion

4.1. 8 vs. 10 h comparisons

Our results suggest that ATC personnel working the 10 h shift schedule do not exhibit any evidence of lower performance on the NIOSH tests across workdays or within workdays than do ATCSs on an 8 h rotating schedule. Regardless of the task (reaction time, digit addition, or grammatical reasoning), none of the between-group differences in performance was statistically significant. This was true for both the reaction time and error measures. Any differences in test performance that were present tended to favor the 10 h ATCSs. This outcome is at contrast with the general findings from investigations concerning 12 h work days. In comparing the 8 and 12 h work days, Rosa and Bonnet (1993), Rosa et al. (1989) and Rosa (1991) all reported that performance on some tests from the NIOSH Fatigue Test Battery was significantly poorer for those employed on 12 h work schedules. However, studies involving the 12 h shift schedule included complete coverage of the 24 h workday, while the ATCSs involved in this 10 h study only covered the afternoon and morning shifts. At present ATCSs at this facility are not assigned to work a 10 h night shift.

4.2. Fatigue test battery sensitivity

Despite the finding of no differences between schedules, the results clearly demonstrate the sensitivity of the tests selected from the NIOSH fatigue test battery to alterations in alertness associated with working a rotating shift schedule. However, the tests included in this study did not appear to be equally sensitive to the effects of either the work day or session within the work day. The choice reaction time measure appeared to be most sensitive to variations in test time and effects of the work week. The night shift, with the associated sleep loss and test times between 10 pm and the early morning hours clearly resulted in slower response times and greater errors. Changes in average values for ATCSs on the night shift were the only occasions where response times declined and errors increased from start to completion of the work day on each of the 3 tests. Subjective ratings of mood were less positive at night and became increasingly so during the course of the work day. Thus, the tests were sensitive to the combined effects of shorter sleep times later in the week and earlier start time for the shift, or a change to the night shift. The extent to which these changes can be attributed to the sleep loss associated with this particular quick-rotating schedule, or to the effects of the circadian rhythm on performance cannot be determined from this study.

The obtained alterations in the fatigue test battery performance measures reflect both the attentional demands of the specific tasks and the overall sensitivity of the component measures to fatigue, but were not necessarily reflected in changes in operational (job) performance. Operational tasks often involve much greater opportunity for analysis and response to critical situations than the tasks presented under these experimental conditions. However, the outcomes do reflect some general decrements in readiness of the human operator to respond that are associated with circadian rhythms and sleep loss resulting from a rotating shift schedule.

4.3. Sleep and mood

The consistent decline in total sleep time from start to completion of the work week reported by the ATCSs who participated in this study appears to be closely related to the general pattern of changes in performance noted on the choice reaction time component of the test battery. The effects were also readily observable in the self-reported positive and negative mood ratings and ratings of feeling "refreshed" following sleep. Shortened sleep times during the first 4 d were nearly identical for ATCSs on both the 8 and 10 h shift schedules. From a high of 8.06 h on the evening prior to the first day of the work week, ATCSs reported progressively fewer hours of sleep each day to 5.77 on day 4. For ATCSs on the 8 h schedule, the combination of the short turnaround and the need to sleep during the daylight hours, resulted in the shortest sleep time prior to the night shift on the final day of the work week (3.57 h).

The number of hours of sleep reported by ATCSs working the 8 h shifts in this study are consistent with those reported in earlier studies by Melton and his colleagues in the 1970s (Melton et al., 1971, 1973, 1975; Saldivar et al., 1977). Melton et al. (1973) commented that the shorter sleep time for ATCSs prior to the night shift may be attributed, in part, to the tendency for some ATCSs to take only a brief nap prior to the night shift so that they will be able to sleep better during the morning following completion of the night shift. These findings are also consistent with outcomes from a recent assessment of controller sleep time on the 2-2-1 schedule at the Miami en route center (Cruz and Della Rocco, 1995), where ATCSs averaged 2.4 h of sleep prior to the night shift. In contrast, controllers at the Miami International Flight Service Station reported slightly longer sleep periods (5.45 h) prior to the night shift (Melton, 1985). These outcomes suggest that individuals develop different strategies relative to the amount of sleep they obtain prior to the night shift.

4.4. *Rapidly rotating shift schedules*

Even though outcomes from this study suggest that the 10 h schedule is acceptable, none of the controllers worked a 10 h schedule involving the night shift. The requirements to provide 24 h ATC coverage, and to have adequate staffing across the work day requires the continuation of various 8 h schedules. Despite some of the obvious advantages of rapidly rotating shift schedules, where employees generally work 2 or fewer nights in succession, research concerning those schedules is not very extensive. Additionally, there is considerable variation in the actual working hours and nature of the proposed rapidly rotating schedules. Despite these facts, Wilkinson (1992), in a brief review of the outcomes from various types of shift schedules, concluded that fixed night systems are superior and should be implemented for night work. The effectiveness of this approach however, is dependent on the willingness of employees to remain on a "night" schedule even during their days off. In reply, Folkard (1992), argued that Wilkinson overestimated the problems associated with rapidly rotating shift systems and that other aspects of shift systems should be taken into account when determining the best shift schedule. In a series of studies, Melton and his colleagues reported that while ATCSs on a 2-2-1 schedule obtain slightly less sleep across the work week than their colleagues on either a 5 d rotating or 5 d fixed-schedule, they did not differ significantly on most of the physiological and biochemical indices of stress (Melton et al., 1971, 1973, 1975; Melton, 1985) or the measures of mood and anxiety (Melton et al., 1971, 1973, 1975). Melton (1985), however, reported that a group of ATCSs employed at the Miami flight service station on the 2-2-1 shift exhibited higher levels of self-reported fatigue prior to the start of their work week than those on a 5 d fixed schedule.

ATCSs who favor the 2-2-1 schedule have consistently reported that this preference is based primarily on the longer number of hours off between work weeks, and that they are required to work only a single night shift. Another social factor associated with the 2-2-1 shift schedule is that a relatively normal amount of sleep and a relatively normal family schedule can be maintained during much of the work week. Ability to maintain a near normal pattern of sleep time is only seriously disrupted just prior to starting the night shift. Additionally, the timing of the change in shifts is such that the staff of ATCSs who handle the typical morning push of air traffic comes from the ATCSs who have just started their work day, rather than those who are completing the night shift.

While there is considerable variation in shift schedule preference among ATCSs, the 2-2-1 schedule has continued to be viewed positively by much of the ATC workforce. This is evidenced, in part, by its continued existence at most ATC facilities for more than 2 decades, as employees, union representatives, and management have conferred regarding the selection of a preferred shift schedule. Anecdotal comments from controllers and facility managers, however, suggest that the percentage of younger controllers preferring the 2-2-1 schedule is greater than that of older controllers. However, as part of an older survey of ATCS job attitudes, Smith (1973) determined that while there was a trend for the preference of the 2-2-1 schedule to diminish with age, it was still the most preferred schedule for older controllers. As the ATC workforce ages over the next decade, continued research will be needed to determine the extent to which older controllers may experience difficulties in coping with the 2-2-1 schedule, and to assess the effectiveness of alternative schedules and fatigue countermeasures that would reduce the negative consequences of working a rotating shift schedule. During the 2 decades of using the 2-2-1 shift schedule at ATC facilities across the US, controllers have provided anecdotal comments concerning difficulties associated with working a rotating shift schedule. However, there is little documented evidence of any significant negative impact on work performance, safety, or overall well-being.

References

Baddeley, A.D., 1968. A 3-minute reasoning task based on grammatical transformation. Psychonomic Science 10, 341–342.

Broach, D., Aul, J.C., 1993. Analysis of the air traffic control specialist (ATCS) occupation using the Position Analysis Questionnaire (PAQ), unpublished manuscript.

Calvasina, E., Boxx, R., 1975. Efficiency of workers on the four day work week. Academy of Management Journal 18, 604–610.

Cruz, C.E., Della Rocco, P.S., 1995. Sleep patterns in Air Traffic Controllers Working Rapidly Rotating Shifts: A Field Study. (DOT/FAA/AM-95/12). Federal Aviation Administration, Washington, DC.

Daniel, J., Potasova, A., 1989. Oral temperature and performance in 8 and 12 h shifts. Ergonomics 32, 689–696.

Duchon, J.C., Smith, T.J., 1993. Extended workdays and safety. International Journal of Industrial Ergonomics 11, 37–49.

Dunham, R.B., Pierce, J.L., Castaneda, M.B., 1987. Alternative work schedules: two quasi-experiments. Personnel Psychology 40, 215–242.

Folkard, S., 1992. Is there a 'best compromise' shift system? Ergonomics 35, 1453–1463.

Goodale, J.G., Aagaard, A.K., 1975. Factors relating to varying reactions to the 4 d workweek. Journal of Applied Psychology 60, 33–38.

Hartman, R.I., Weaver, K.M., 1977. Four factors influencing conversion to four-day work week. Human Resource Management 16, 24–27.

Hodge, B.J., Tellier, R.D., 1975. Employee reactions to four-day work week. California Management Review 18, 25–30.

Johnson, L.C., Naitoh, P., 1974. The operational consequences of sleep deprivation and sleep deficit. AGARDograph AF-193, NATO Advisory Group for Aerospace Research and Development. London, England.

Kenny, M.T., 1974. Public employee attitudes toward the four-day work week. Public Personnel Management 3, 159–161.

Latack, J.C., Foster, L.W., 1985. Implementation of compressed work schedules: Participation and job redesign as critical factors for employee acceptance. Personnel Psychology 38, 75–93.

Lewis, P.M., Swaim, D.J., 1988. Effect of a 12-hour/day shift on performance. In: Hagen, E.W. (Ed.) Proceedings of the Fourth Institute of Electrical and Electronics Engineers Conference on Human Factors and Power Plants. Institute of Electrical and Electronics Engineers, New York, pp. 513–516.

Melton, C.E. 1985. Physiological responses to unvarying (steady) and 2-2-1 shifts: Miami International Flight Service Station, (DOT/FAA/AM-85/2). Federal Aviation Administration, Washington, DC.

Melton, C.E., McKenzie, J.M., Polis, B.D., Funkhouser, G.E., Iampietro, P.F., 1971. Physiological responses in air traffic control personnel: O'Hare Tower, FAA/AM-71-2. Federal Aviation Administration, Washington, DC.

Melton, C.E., Smith, R.C., McKenzie, J.M., Saldivar, J.T., Hoffman, S.M., Fowler, P.R., 1975. Stress in air traffic controllers: comparison of two air route traffic control centers on different shift rotation patterns, FAA/AM-75-7. Federal Aviation Administration, Washington, DC.

Melton, C.E., McKenzie, J.M., Smith, R.C., Polis, B.D., Higgins, E.A., Hoffman, S.M., Funkhouser, G.E., Saldivar, J.T., 1973. Physiological, biochemical, and psychological responses to air traffic control personnel: comparisons of the 5-day and 2-2-1 shift rotation patterns, FAA/AM-73-22. Federal Aviation Administration, Washington, DC.

Mills, M.E., Arnold, B., Wood, C.M., 1983. Core-12: A controlled study of the impact of 12-hour scheduling. Nursing Research 32, 356–361.

Myers, J.L., 1979. Fundamentals of Experimental Design. Allyn and Bacon, New York.

Peacock, B., Glube, R., Miller, M., Clune, P., 1983. Police officers' responses to 8 and 12 h shift schedules. Ergonomics 26, 479–493.

Rosa, R.R., 1991. Performance, alertness and sleep after 3–5 years of 12 h shifts: a follow-up study. Work and Stress 5, 107–116.

Rosa, R.R., Bonnet, M.H., 1993. Performance and alertness on 8 and 12 h rotating shifts at a natural gas utility. Ergonomics 36, 1177–1193.

Rosa, R.R., Colligan, M.J., 1988. Long workdays versus restdays: assessing fatigue and alertness with a portable performance battery. Human Factors 30, 305–317.

Rosa, R.R., Colligan, M.J., Lewis, P., 1989. Extended workdays: Effects of 8-hour and 12-hour rotating shift schedules on performance, subjective alertness, sleep patterns, and psycho-social variables. Work and Stress 3, 21–32.

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

Saldivar, J.T., Hoffman, S.M., Melton, C.E., 1977. Sleep in air traffic controllers, FAA/AM-77-5. Federal Aviation Administration, Washington, DC.

Smith, R.C., 1973. Job attitudes of air traffic controllers: A comparison of three air traffic specialties, FAA/AM-73-2. Federal Aviation Administration, Washington, DC.

Steele, J.L., Poor, R., 1970. Work and leisure: The reactions of people at 4-day firms. In: Poor, R. (Ed.), 4 Days, 40 Hours. Bursk and Poor, Cambridge, MA, pp. 105–122.

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

Wheeler, M.E., 1970. Four days, forty hours: a case study. California Management Review 17, 74–81.

Wilkinson, R.T., 1992. How fast should the night shift rotate? Ergonomics 35, 1425–1446.

Williams, H.L., Lubin, A., 1966. Speeded addition and sleep loss. Journal of Experimental Psychology 73, 313–317.