# The Ability to Self-Monitor Performance During a Week of Simulated Night Shifts

Jillian Dorrian, PhD; Nicole Lamond, PhD; Alexandra L. Holmes, PhD; Helen J. Burgess, PhD; Gregory D. Roach, PhD; Adam Fletcher, PhD; Drew Dawson, PhD

*The Centre for Sleep Research, The University of South Australia*

**Study Objectives:** Research has indicated that individuals are able to accurately monitor the performance decrements they experience during unitary periods of acute sleep deprivation. The aim of the current study was to investigate the ability to self-monitor performance during a week of simulated night shifts.

**Design:** Subjects completed 7 consecutive 8-hour night shifts (11 pm-7 am).

**Setting:** University sleep laboratory

**Subjects:** Fifteen young (7 men, 8 women, 19-25 years) healthy volunteers

**Interventions:** During the night shifts, performance was measured hourly on 4 performance parameters: psychomotor vigilance test (PVT), tracking, and grammatical reasoning (GRG) accuracy and response latency. Before and after each test, subjects completed visual analogue scales, which required them to rate their alertness and their performance speed and/or accuracy.

**Results:** Analysis indicated that GRG response latency and tracking were significantly impaired ($P<0.05$) during the first 2 shifts only. The PVT performance displayed consistent impairment, with significant ($P<0.05$) declines during all but the final shift. The pattern of deterioration in subjective ratings of alertness was similar to that of the PVT data. Correlations between subjective alertness and self-ratings of performance were significant ($P<0.01$) for all parameters ($r=0.39$-$0.69$). Significant ($P<0.05$) correlations were found across the week between pretest performance ratings and actual performance for all parameters except GRG accuracy ($r=0.29$-$0.58$) and between posttest ratings and actual performance for all parameters ($r=0.52$-$0.75$). Correlations between pretest ratings and actual performance were also conducted separately for each shift. Highest correlations were found during the first shift, with r-values that were low for GRG accuracy ($r=0.32$) and GRG response latency ($r=0.20$), moderate for tracking ($r=0.41$), and high for PVT ($r=0.82$). In general, lower correlations were found later in the week.

**Conclusions:** Overall, results indicate that individuals have only a moderate ability to predict performance impairment during a week of night shifts. It is likely that performance ratings are based, at least to a certain extent, on subjective alertness levels. Furthermore, it seems that rating accuracy is improved on tasks providing performance feedback, such as the PVT. Finally, it appears that after testing, individuals have a more accurate perception of their performance.

**Key Words:** Self-monitoring, performance, shift work, fatigue

**Citation:** Dorrian J; Lamond N; Holmes AL et al. The ability to self-monitor performance during a week of simulated night shifts. *SLEEP* 2003; 26(7):871-7.

## INTRODUCTION

THE DETRIMENTAL EFFECTS OF SHIFTWORK, IN PARTICULAR NIGHT WORK, HAVE BEEN WELL DOCUMENTED.[1-4] Field and laboratory studies indicate that, in general, workers are sleepier at night[5,6] and that worker performance is slower and less accurate.[7-9] As a consequence, night work is often associated with decreased efficiency, lowered productivity, and increased risk of accidents.[3,10]

It has been postulated that individual awareness of performance impairment may trigger compensatory responses that reduce or eliminate potential error.[11,12] If this is so, accurate knowledge of performance decrements may serve to reduce error and accident risk in the workplace and maintain profitability. In many operational situations, immediate information concerning performance efficacy is not accessible to the worker. In such situations, the only mechanism for performance feedback available to an individual is self-assessment. For example, in the case of long-haul truck drivers who travel alone for extended periods, introspection is often the only performance-assessment method possible. However, it has been suggested that sleep loss and fatigue impair the capacity to introspect[13] and reduce the metacognitive ability to realistically assess impairment.[14] It is possible, then, that insight into performance ability may deteriorate with the increasing fatigue commonly associated with night work.

Contrary to the suggestion that sleep loss and fatigue affect the ability to self-assess impairment, studies have indicated that fatigued individuals are usually aware of reduced performance capability and are, therefore, able to give accurate self-ratings of performance.[11,15,16] In line with these findings, a recent study by our research group demonstrated that, following a period of acute sleep loss, fatigued subjects were able to accurately rate deterioration of their performance on several neurobehavioral tasks. Furthermore, as equivalent ratings were given for all measures, despite the fact that not all of the performance measures were sensitive to the effects of fatigue, it was concluded that self-ratings of performance were global rather than test specific. Finally, it was apparent that the subjective ratings were, at least to some extent, mediated by the individual's level of alertness.[17]

To date, studies investigating self-awareness of impairment have solely focused on the effects of single periods of sleep deprivation. Since shiftworkers typically work several consecutive night shifts, findings of such studies are limited in workplace generalizability. Research has shown that while day workers are typically awake for 1 to 2 hours before commencing work, night-shift workers are usually awake between 10 and 16 hours.[18] Furthermore, research has shown that 50% of shiftworkers typically spend at least 24 hours awake on the first night of a roster.[19] As such, shift workers often experience acute sleep deprivation on the first of a series of night shifts. In addition, due to the subsequent desynchronization of the sleep-wake cycle and circadian system, the daytime sleep of night workers is of shorter duration and poorer quality than the night sleep of day workers.[3,20] Therefore, shiftworker fatigue is usually due to the combination of cumulative sleep loss and circadian desynchrony, rather than simply acute sleep loss.

**Disclosure Statement**

Research supported by the Australian Research Council

**Submitted for publication August 2002**

**Accepted for publication May 2003**

Address correspondence to: Jillian Dorrian, The Centre for Sleep Research, 5th Floor Basil Hetzel Institute, The Queen Elizabeth Hospital, Woodville Rd, Woodville SA 5011, Australia; Tel: +61 8 8222 6624; Fax: +61 8 8222 6623; E-mail: jill.dorrian@unisa.edu.au

Therefore, the aim of the current study was to investigate the accuracy of self-ratings of performance over a week of simulated night shifts.

## METHODS

### Subjects

Fifteen subjects (7 men, 8 women), with a mean age of 21.9 years (SD, 2.7 years), participated in the study. For screening purposes, volunteers were required to complete a general health questionnaire. Subjects who had a current health problem or a history of psychiatric or sleep disorders were excluded, as were subjects who smoked cigarettes, consumed large amounts of caffeine (>350 mg/day) or were taking any medication known to effect sleep or performance. All women in the study were taking an oral contraceptive pill. Subjects were not regular nappers and had never participated in shift work, nor had they traveled transmeridian in the month prior to the study.

### Procedure

All testing was carried out at the Centre for Sleep Research, at the Queen Elizabeth Hospital. Ethics approval for the project was granted by the Queen Elizabeth Hospital Ethics Committee.

Subjects attended the laboratory for 9 consecutive nights. The first 2 nights consisted of an adaptation night and a baseline night. These were directly followed by 7 consecutive nights of simulated shiftwork. For the adaptation and baseline nights, subjects arrived at the sleep laboratory at 5:00 PM. Using prestudy 7-day sleep diaries, average bedtimes for each subject were determined and assigned as the bedtime for these nights. Subjects were instructed to sleep until they woke naturally, after which they were permitted to leave the laboratory for the day.

On the third night, subjects returned by 7:00 PM for the first of a series of 7 simulated nightshifts. During each shift (beginning at 11:00 PM and ending at 7:00 AM), subjects completed hourly performance testing sessions of approximately 20 minutes duration. Between testing sessions, subjects were free to read, watch television, play games, and talk quietly. Careful monitoring by researchers ensured wakefulness over the testing period.

At 7:00 AM, following the final testing session, subjects were accompanied outside for a 15-minute period in order to expose them to the sunlight a night-shift worker would typically be exposed to when traveling home from work. Subjects then retired at approximately 8:00 AM and were instructed to sleep until they awoke naturally. However, subjects were awakened by the experimenter if they were still asleep at 7:00 PM (this occurred on 2 occasions). After waking, subjects were free to leave the laboratory until 7:00 PM. Subjects were instructed that they were not allowed to nap during this free period. In order to verify that this was the case, subjects were monitored using wrist actigraphs (data reported elsewhere).

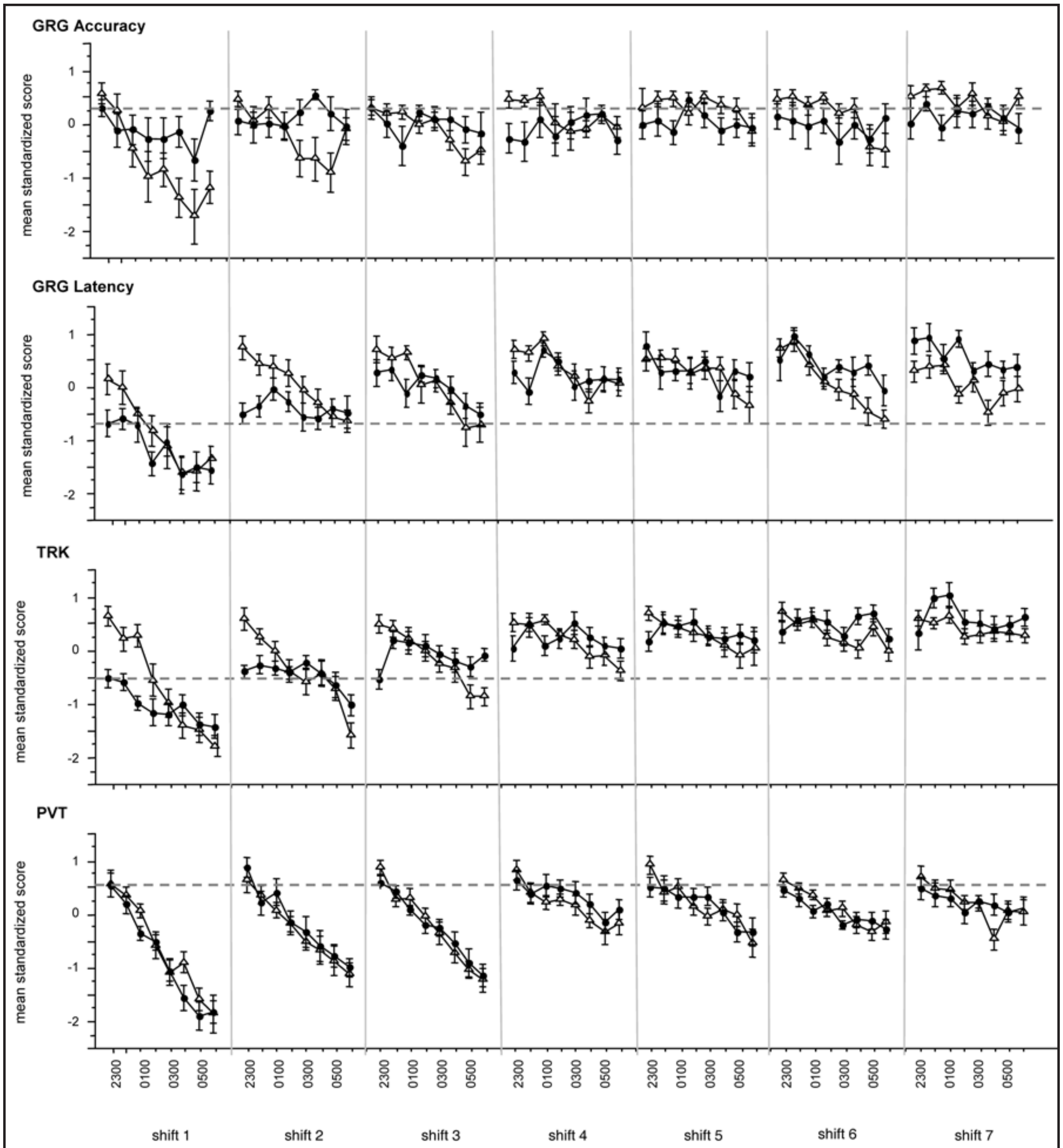


**Figure 1**—Actual performance (filled circles), and pre-test performance ratings (open triangles) for each performance parameter. Dotted lines are positioned at the mean performance level during the first testing session of the week.

### Neurobehavioral Tests

Neurobehavioral performance was measured using tests of unpredictable tracking, grammatical reasoning (GRG), and psychomotor vigilance (PVT). In order to minimize practice effects, subjects attended two 5-hour training sessions during the week prior to the experimental condition. During these sessions, subjects were trained on each of the performance tasks until performance reached a plateau.

The tracking and GRG tasks were from a standardized computer-based test battery (Worksafe, Australia), based on a standard information-processing model,[21] and have been described in detail elsewhere.[17,22] The measures used in this report include (1) tracking accuracy, (2) GRG accuracy, and (3) GRG response latency. In addition, response time on a 10-minute PVT task was evaluated.[23,24]

During test sessions, subjects were seated in an isolated room, free of distraction, and were instructed to complete each task once (tasks were presented in a random order to prevent order effects). Each test session lasted approximately 20 minutes. Because knowledge of results can influence performance levels,[25] subjects received no verbal feedback from the experimenters during the study. However, reaction times are displayed on the stimulus screen for each response on the PVT, providing concurrent visual performance feedback for this task.

### Self-rating Scales

Self-rating scales were used to record pretest and posttest subjective performance ratings. The scales consisted of linear, nonnumeric, 100-millimeter visual analogue scales (VAS), anchored at each end with semantic opposites. The scales were administered separately, immediately before and after each test within the battery.

Questions for the self-rating scales fit the format; "How accurately/quickly do you think you will perform/performed on the next/previous …… task?" Scales for the tracking task were anchored with "0% of the time on target" at one end and "100% of the time on target" on the other. Scales for the accuracy component of the GRG task were marked "none correct" at one end, and "all correct" on the other. Scales for GRG and PVT response latency were anchored with "extremely slowly" at one end and "extremely quickly" at the other.

In addition, subjects were asked to rate their alertness before each testing session. Subjects were asked, "How alert do you feel?" Scales were anchored with "struggling to remain awake,"' and "extremely alert and wide awake." Subjects were instructed to complete the ratings relative to their perception of their own average alertness and performance levels from the training session.

### Sleep-Wake Activity

Polysomnographic (PSG) data were collected during all of the sleep periods. Sleep-wake state was assessed using a standard electroencephalographic montage (C3/A2, C4/A1), electrooculogram, and electromyogram in accordance with the International 10/20 System.[26] Both electroencephalographic signals were sampled within a 0.33- to 70-Hz bandwidth, digitized at 250 Hz and filtered with a 50-Hz notch filter. This was done using 2 sleep recording systems: Compumedics 10-20 system (Melbourne, Australia) and Medilog MPA-2 sleep analysis system (Oxford Medical Ltd, Oxton, England). The PSG data were double scored according to the standard criteria of Rechtschaffen and Kales.[26] In the case of scorer discrepancies, a third independent scorer made a final decision as to sleep stage. For the purposes of this report, measures derived from the PSG data included total sleep time (TST), sleep efficiency for each sleep period (calculated as the TST/ sleep period time x 100) and cumulative sleep debt (calculated as the sum of the difference [minutes] between the baseline sleep and each subsequent day sleep).

**Table 1—**Summary of 2-way ANOVA for shift (1-7) and trial (1-8) for actual performance, subjective alertness, and pretest and posttest performance ratings.

| | Shift | | Trial | | Shift * Trial | |
|---|---|---|---|---|---|---|
| | $F_{(6, 588)}$ | *P* | $F_{(7, 588)}$ | *P* | $F_{(42, 588)}$ | *P* |
| Actual | | | | | | |
| GRG accuracy | 0.884 | NS | 0.546 | NS | 0.608 | NS |
| GRG latency | 25.704 | 0.0001 | 5.184 | 0.0016 | 1.185 | NS |
| TRK | 28.631 | 0.0001 | 4.047 | 0.0071 | 1.342 | NS |
| PVT | 9.291 | 0.0001 | 33.826 | 0.0001 | 4.004 | 0.0002 |
| Alertness | 15.239 | 0.0001 | 28.938 | 0.0001 | 3.768 | 0.0007 |
| Pretest | | | | | | |
| GRG accuracy | 7.474 | 0.0009 | 12.166 | 0.0001 | 1.969 | NS |
| GRG latency | 7.356 | 0.0002 | 22.252 | 0.0001 | 0.0774 | NS |
| TRK | 14.249 | 0.0001 | 25.759 | 0.0001 | 3.581 | 0.0005 |
| PVT | 7.976 | 0.0001 | 52.078 | 0.0001 | 2.730 | 0.0049 |
| Posttest | | | | | | |
| GRG accuracy | 2.415 | NS | 1.217 | NS | 1.036 | NS |
| GRG latency | 2.071 | NS | 4.328 | 0.0056 | 0.571 | NS |
| TRK | 16.0005 | 0.0001 | 9.527 | 0.0001 | 2.398 | 0.0099 |
| PVT | 14.102 | 0.0001 | 20.599 | 0.0001 | 2.491 | 0.0088 |

*P* values corrected by Greenhouse-Geisser Epsilon. GRG, grammatical reasoning; TRK, tracking; PVT, psychomotor vigilance test

### Statistical Analysis

According to standard methodology, PVT data were transformed to 1 per reaction time to correct for proportionality between the mean and SD.[24,27] To control for interindividual variability on neurobehavioral performance, and to allow comparison of predicted and actual performance scores, all scores for each subject were converted to z-scores.

Evaluation of systematic changes in subjective alertness and predicted and actual performance parameters were assessed by 2-way repeated measures analysis of variance (ANOVA), for shift (1-7) and trials (1-8). Individual repeated measures ANOVA for trials (1-8) within each shift were also conducted. In addition, repeated measures ANOVA for days (1-6) were used to assess changes in sleep parameters. Significance levels were corrected to produce more conservative degrees of freedom by Greenhouse-Geisser Epsilon.

Time series correlation (TSC) coefficients were calculated for each participant -3 to +3 time lags (1 lag=1 hour [or 1 trial]), for comparisons between actual performance, pretest predicted performance, posttest performance estimates, and pretest and posttest subjective alertness. For all TSCs, r-values were found to be highest at a time lag of 0, indicating no clear lag effect. As such, r-values will be reported at time-lag 0 only. Since distributions of r-values are highly skewed, an average r across all subjects for each test was obtained using Fisher's r-z transformation. Correlation analyses were conducted both for each shift separately and across the entire week of shifts. It must be noted, however, that correlations across the week involved both within and between subjects factors, and there is therefore an increased risk of Type 1 error.

## RESULTS

### Actual Performance

Two-way ANOVA revealed a significant ($P$<0.001) effect of shift (1-7) and trials (1-8) for GRG response latency, tracking, and PVT (see Table 1 and Figure 1). Table 2 shows the results of individual ANOVA for performance within each shift. Scores significantly ($P$<0.05) declined during shift 1 only for GRG response latency and during shifts 1 and 2 for the tracking task. The PVT task showed significant deterioration for shifts 1 to 6. The PVT was the only task for which a significant ($P$<0.0001) interaction effect of shift and trials was observed. Declines in PVT scores within each shift were less marked as the week progressed. No significant variation was observed for the GRG accuracy (see Table 1).

**Table 2—**Summary of ANOVA for trials within each shift for performance on GRG latency, tracking, and PVT and for subjective alertness ratings.

| Shift | GRG latency | | TRK | | PVT | | Alertness | |
|---|---|---|---|---|---|---|---|---|
| | $F_{(7, 98)}$ | *P* | $F_{(7, 98)}$ | *P* | $F_{(7, 98)}$ | *P* | $F_{(7,98)}$ | *P* |
| 1 | 4.346 | 0.0016 | 3.191 | 0.0191 | 30.322 | 0.0001 | 13.92 | 0.0001 |
| 2 | 0.944 | NS | 2.752 | 0.0363 | 12.391 | 0.0001 | 11.313 | 0.0001 |
| 3 | 1.964 | NS | 2.143 | NS | 13.922 | 0.0001 | 17.022 | 0.0001 |
| 4 | 1.462 | NS | 1.284 | NS | 2.569 | 0.0493 | 9.033 | 0.0001 |
| 5 | 1.215 | NS | 0.584 | NS | 4.098 | 0.0282 | 3.632 | 0.0289 |
| 6 | 1.611 | NS | 1.153 | NS | 4.174 | 0.0061 | 5.448 | 0.0050 |
| 7 | 1.831 | NS | 1.870 | NS | 1.087 | NS | 1.144 | NS |

*P* values corrected by Greenhouse-Geisser Epsilon. GRG, grammatical reasoning; TRK, tracking; PVT, psychomotor vigilance test

### Subjective Alertness

Subjective alertness ratings are illustrated in Figure 2. Two-way ANOVA (see Table 1) revealed significant main effects of shift and trial for subjective alertness ratings ($P<0.0001$). Individual ANOVA for trials within each shift (see Table 2) revealed significant declines in alertness during shifts 1 to 6 ($P<0.0001$). In addition, a significant ($P<0.0001$) interaction effect of shift and trial was found such that declines in alertness became less severe over the week.

**Table 3—** Summary of ANOVA for trials within each shift for pretest performance ratings for each performance measure.

| Shift | GRG accuracy | | GRG latency | | TRK | | PVT | |
|---|---|---|---|---|---|---|---|---|
| | $F_{(7, 98)}$ | $P$ | $F_{(7, 98)}$ | $P$ | $F_{(7, 98)}$ | $P$ | $F_{(7, 98)}$ | $P$ |
| 1 | 5.341 | 0.0016 | 7.032 | 0.0001 | 20.331 | 0.0001 | 22.497 | 0.0001 |
| 2 | 3.123 | 0.0261 | 5.837 | 0.0001 | 10.998 | 0.0001 | 11.692 | 0.0001 |
| 3 | 5.560 | 0.0025 | 6.346 | 0.0006 | 7.549 | 0.0002 | 21.048 | 0.0001 |
| 4 | 1.789 | NS | 4.542 | 0.0018 | 5.652 | 0.0001 | 4.379 | 0.0120 |
| 5 | 1.226 | NS | 2.474 | NS | 1.802 | NS | 5.294 | 0.0029 |
| 6 | 4.946 | 0.0054 | 7.698 | 0.0001 | 3.404 | 0.0165 | 7.610 | 0.0002 |
| 7 | 2.213 | NS | 3.123 | 0.0174 | 1.762 | NS | 5.537 | 0.0008 |

*P* values corrected by Greenhouse-Geisser Epsilon. GRG, grammatical reasoning; TRK, tracking; PVT, psychomotor vigilance test

**Table 4—**Summary of ANOVA for trials within each shift for posttest performance ratings for GRG latency, tracking, and PVT.

| Shift | GRG latency | | TRK | | PVT | |
|---|---|---|---|---|---|---|
| | $F_{(7, 98)}$ | $P$ | $F_{(7, 98)}$ | $P$ | $F_{(7, 98)}$ | $P$ |
| 1 | 2.008 | NS | 6.485 | 0.0001 | 13.425 | 0.0001 |
| 2 | 2.323 | NS | 5.018 | 0.0014 | 11.077 | 0.0001 |
| 3 | 1.143 | NS | 2.470 | NS | 7.460 | 0.0001 |
| 4 | 1.069 | NS | 4.028 | NS | 4.052 | 0.0053 |
| 5 | 1.185 | NS | 1.875 | NS | 1.859 | NS |
| 6 | 1.870 | NS | 2.747 | NS | 2.617 | NS |
| 7 | 1.190 | NS | 0.422 | NS | 1.082 | NS |

*P* values corrected by Greenhouse-Geisser Epsilon. GRG, grammatical reasoning; TRK, tracking; PVT, psychomotor vigilance test

**Table 5—**Results of time-series correlations between actual performance, pretest ratings, posttest ratings and pretest subjective alertness ratings. R-values are reported at time-lag 0.

| Performance Measure | Pretest & Actual | Pretest & Alertness | Actual & Posttest | Posttest & Alertness |
|---|---|---|---|---|
| GRG accuracy | 0.09 | **0.48 | **0.68 | 0.14 |
| GRG latency | *0.29 | **0.61 | **0.52 | **0.37 |
| TRK | **0.42 | **0.69 | **0.53 | **0.53 |
| PVT | **0.58 | **0.39 | **0.75 | *0.29 |

*$P<0.05$, **$P<0.01$ GRG, grammatical reasoning; TRK, tracking; PVT, psychomotor vigilance test

### Pretest Performance Ratings

Significant ($P<0.0001$) main effects of shift and trial were observed for pretest performance ratings for all 4 performance parameters (Table 1 and Figure 1). Results of individual ANOVA for trials within each shift for pretest ratings are displayed in Table 3. Ratings predicted that performance on GRG accuracy would significantly ($P<0.05$) decline for shifts 1, 2, 3, and 6. Ratings predicted significant ($P<0.05$) impairment in GRG response-latency performance during all shifts except shift 5 and in tracking during all shifts except 5 and 7. The PVT pretest ratings predicted significant ($P<0.05$) performance impairment during all shifts. A significant ($P<0.01$) interaction effect of shift and trial was found for pretest ratings for tracking and PVT. For these measures, pretest ratings predicted a reduction in performance impairment with increasing night shifts.

### Posttest Performance Ratings

A significant ($P<0.0001$) main effect of shift was found for posttest performance ratings for all parameters (Table 1). The effect of trial was significant for all parameters except GRG accuracy ($P<0.0001$). Results of individual ANOVA for trials within each shift for posttest ratings are displayed in Table 4. Despite the significant effect of trial found for GRG response latency, individual ANOVA revealed no significant declines during any of the 7 shifts. However, variation during shift 2 was approaching significance ($F_{(7, 98)}=2.323$, $P_{(uncorrected)}=0.030$, $P_{(G\text{-}G\ correction)}=0.0756$). Posttest ratings indicated significant ($P<0.01$) impairment for tracking during shifts 1 and 2 and for PVT during shifts 1 to 4. A significant ($P<0.05$) interaction effect of shift and trial was found for posttest ratings for tracking and PVT.

### Correlation Analysis

As can be seen in Table 5 (also illustrated in Figure 2), TSC between predicted and actual performance across the week revealed significant ($P<0.05$-$0.01$) low to moderate r-values for all performance parameters except GRG accuracy.

Table 6 displays the results of TSC analysis between predicted and actual performance for each individual shift. Correlations between predicted and actual performance for each shift were low at best for GRG accuracy and response latency. Moderate correlations were found for shifts 1 and 2 on the tracking measure. For the PVT, correlations were moderate to high for shifts 1 to 5 and low for shifts 6 and 7. The only statistically significant correlation was found for the PVT for shift 1. In general, correlations were highest on night 1, becoming lower later in the week.

As seen in Table 5, moderate significant correlations ($P<0.01$) between pretest subjective alertness and predicted performance for all parameters were found.

Moderate to high significant ($P<0.01$) correlations were found between posttest ratings and actual performance for all performance parameters. Results of TSC between posttest ratings and subjective alertness yielded r-values that were low to moderate and were significant ($P<0.01$) for all performance parameters except the GRG accuracy.

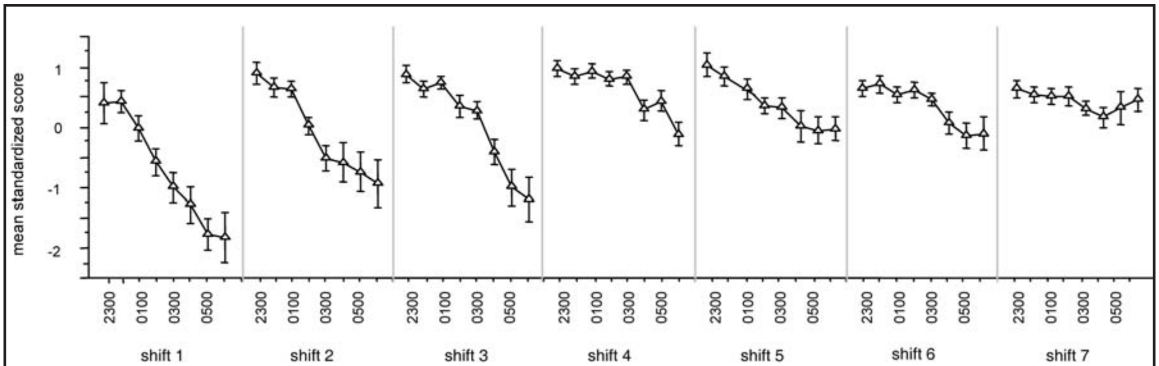

**Figure 2**—Subjective alertness ratings.

### Sleep Duration and Quality

Participants averaged 7.52 hours (SD = 0.61) of sleep during the baseline night. Paired *t*-test results indicated that this was not significantly different from the average sleep duration (mean = 7.64 hours, SD =

0.96) in the week prior to the study, as calculated from sleep diaries ($t_{14}$=-0.586, $P$>0.05). During the 6 consecutive daytime sleeps, in order from sleep 1 to 6, participants averaged 7.02 hours (SD = 1.48), 6.47 hours (SD = 1.30), 7.19 hours (SD = 1.51), 6.99 hours (SD = 1.44), 7.00 hours (SD = 1.20), and 7.24 hours (SD = 1.65), respectively. Consequently, each daytime sleep period was reduced by 35 minutes (on average), relative to the baseline sleep. Analyses indicated that TST did not vary significantly across the 6 daytime sleep periods ($F_{(5, 70)}$=1.15, $P$>0.05). There were also no significant differences between TST on the baseline night and any of the subsequent daytime sleeps ($F_{(5, 70)}$=1.67, $P$>0.05)(see Figure 3a).

Sleep efficiency showed an obvious trend towards greater efficiency across the consecutive daytime sleeps (see Figure 3b). However, the pattern was not quite statistically significant ($F_{(5, 70)}$=2.23, $P$=0.06) nor did efficiency during any of the daytime sleeps significantly differ from the baseline night ($F_{(5, 70)}$=2.53, $P$>0.05).

As displayed in Figure 3c, the cumulative sleep debt significantly increased ($F_{(5, 70)}$=5.10, $P$=0.035) across the week. Prior to the final night shift, there was an average cumulative sleep debt of 3.39 (SD = 5.85) hours.

## DISCUSSION

### Performance

Despite initial impairment, adaptation occurred during the week of simulated night shifts such that performance was reasonably well maintained. The GRG accuracy remained relatively unaffected, while GRG response latency and tracking were only impaired during the first 2 shifts. The PVT was the only task displaying consistent impairment, with declines during all but the final shift. Interestingly, the pattern of deterioration in subjective ratings of alertness was similar to that of the PVT data.

In general, performance during the first night shift in the current study was similar to that observed in a previous study by our research group investigating performance during 1 night awake.[22] In both studies, scores on the accuracy component of the GRG task remained reasonably stable, while declines were observed for the response latency component. This may be due to the instruction given to subjects during training to concentrate on accuracy at the expense of speed or it may reflect a "speed/accuracy trade-off," which has been reported in studies using self-paced tasks of a similar nature.[28-30] In addition, both studies found tracking, vigilance, and subjective alertness to be significantly impaired.

Previous research indicates that night-shift workers experience daytime sleep of reduced duration and quality[3,20] and that this can result in increasing performance impairment as a function of nights on shift.[20] As such, we expected to observe continuing deterioration in alertness and performance over the week. However, this was not the case for GRG reasoning or tracking performance. A necessary consideration when interpreting these results is the potential masking influence of task learning. Indeed, this is a common criticism of studies involving a restricted sleep regime with performance testing over multiple days.[23] Subjects attended two 5-hour training days, during which they practiced all tests until their performance reached a plateau. To achieve this on the GRG reasoning and tracking tasks, subjects completed at least 30 trials.

**Table 6**—Results of time-series correlations between actual performance and pretest performance ratings. R-values are reported at time-lag 0.

| Performance Measure | Shift 1 | Shift 2 | Shift 3 | Shift 4 | Shift 5 | Shift 6 | Shift 7 |
|---|---|---|---|---|---|---|---|
| GRG accuracy | 0.32 | -0.06 | 0.12 | -0.12 | -0.09 | -0.04 | -0.03 |
| GRG latency | 0.20 | 0.15 | 0.34 | 0.03 | -0.03 | 0.30 | 0.28 |
| TRK | 0.41 | 0.39 | 0.11 | 0.01 | 0.17 | 0.12 | 0.09 |
| PVT | *0.80 | 0.63 | 0.62 | 0.55 | 0.51 | 0.38 | 0.20 |

*$P$<0.05. GRG, grammatical reasoning; TRK, tracking; PVT, psychomotor vigilance test

Despite the task training, visual inspection of Figure 1 suggests that it is likely that further learning occurred over the week on these tasks. Dotted lines on this figure indicate the mean performance score on each task during the first testing session of the week. Findings from our previous study investigating performance during 1 night awake[22] indicated that performance remained relatively stable during the first 17 hours of sustained wakefulness (ie, from 8:00 AM on day 1 to 1:00 AM on day 2). During the current study, subjects awoke from their baseline sleep between 8:24 AM and 10:45 AM. As such, during the first testing session (11:00 PM), subjects were experiencing between 12.25 and 14.5 hours of wakefulness. It could therefore be expected that subjects would be performing at near-optimum levels during this first session. It is clear that maximum performance scores on GRG response latency and tracking improved well beyond this level over the week, suggestive of a learning effect. This is consistent with previous findings indicating that reaching asymptotic levels of performance during training is not necessarily adequate if performance testing is to occur periodically over numerous days or weeks.[31]

In contrast, performance on the PVT and GRG accuracy did not greatly improve beyond initial test performance (see Figure 1). As previously mentioned, subjects were instructed to try to achieve 100% accuracy on the GRG task at the expense of speed. In addition, the PVT is associated with minimal learning effects,[2,23,24] with a learning curve of only 1 to 3 trials.[32] As such, it is unlikely that learning effects had an influence on the accuracy component of the GRG task or the PVT.

The PVT was the only task that was consistently impaired during the week, with significant declines during shifts 1 to 6. This is probably due to the relative length, simplicity, and monotony of the task. Indeed, research has shown that tasks that are long and lacking in interest or complexity may be more easily affected by sleep loss.[33] Nevertheless,

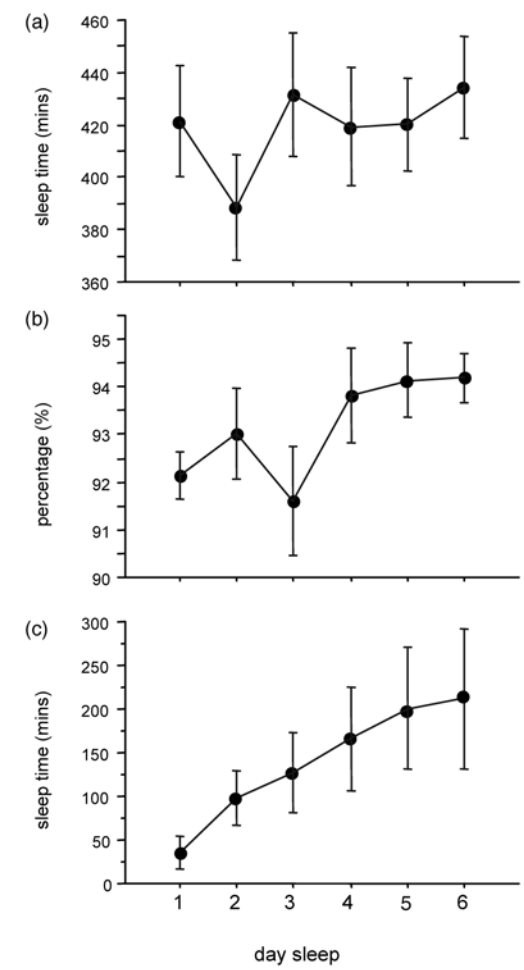


**Figure 3**—TST (a), sleep efficiency (b), and cumulative sleep debt (c) over the six days.

PVT performance improved over the week, such that the decline in performance was less severe as the week progressed. This pattern, paralleled by subjective alertness ratings, suggests circadian adaptation to working nights and subsequent improvement in maintenance of performance efficacy for this measure. In line with this suggestion, research has shown that adjustment to night work takes place over a series of shifts, as the worker's internal biologic clock phase shifts.[34] Moreover, it has been demonstrated that performance follows a circadian rhythm that adjusts, albeit slowly, to night work.[35]

It is important to note that subjects in this study may not have experienced the same level of sleep loss as a 'real' shiftworker attempting to sleep at home. During the study, the laboratory was very dark and quiet, with an absence of the environmental factors (eg, noise, light, phone calls) that may interfere with daytime sleep in a home environment. Furthermore, since this "week of night shifts," was a 1-time occurrence in the lives of our non-shiftworking subjects, the drive to see family and friends during the week in the laboratory was potentially not as great as it would be for a shiftworker who was consistently working night shifts. Therefore, subjects did not tend to sacrifice sleep time for social time while in the laboratory, a situation that would forseeably curtail sleep in the 'real world.' For these reasons, laboratory studies frequently show better sleep than would be expected at home.[36] Indeed, in the current study, neither TST nor sleep efficiency was significantly affected while working a week of night shifts. Moreover, sleep efficiency showed a tendency to improve as the week progressed.

Taken together, results indicate a differential effect of simulated night work on the 4 performance parameters. This is consistent with previous studies finding that while performance on some tests may remain unaffected during the night, others show significant impairment.[22,37] Indeed, such results highlight the importance of multiple approaches to performance testing, using a battery of tasks.[29,38]

### Self-ratings of Performance

Analysis of the week as a whole revealed a moderate association between predicted and actual performance for all performance parameters except GRG accuracy. Interestingly, this was the only parameter displaying no significant effect of fatigue. This is consistent with a previous study by our research group[17] and indicates that subjects had a moderate ability to predict changes in their performance over the week.

Also consistent with our previous findings,[17] predicted performance closely paralleled the pattern of subjective alertness, such that a strong relationship was found between the 2 measures. Although correlation analysis can in no way determine direction of causality, we suggest that pretest performance ratings were affected to some extent by alertness levels. Intuitively, subjective alertness seems a likely cue for performance judgement. Indeed, as they became more fatigued and less alert, subjects typically expected that their performance would deteriorate. Furthermore, as the week progressed and declines in alertness became less marked, this was reflected in the pretest performance ratings.

The strongest relationship between predicted and actual performance was observed for the PVT. Since this is the only test that provides subjects with performance feedback, the finding that scores on this test were best predicted is easily understood. During the 10-minute test, reaction times for each required response were displayed on the stimulus screen. In this way, subjects were provided with direct information regarding their performance and were thus in a better position to predict future performance. Indeed, research has indicated that people who predict performance using information about their own past behavior are more accurate than those using more general information.[39] Moreover, the lowest correlation between predicted performance and subjective alertness was observed for the PVT. Taken together, these results suggest that when subjects had more direct access to information regarding their performance, subjective alertness mediated their pretest performance predictions to a lesser degree. In other words, it seems that subjective alertness provides at least a partial basis for predictions of future performance, in the absence of more direct performance feedback.

As has been found previously,[17,40] posttest self-ratings were better predictors of actual performance than were pretest self-ratings. While pretest and posttest judgments appear to be mediated to some extent by subjective alertness, a stronger relationship was found between pretest ratings and alertness. As such, it is probable that posttest judgments were based to a greater degree on actual performance. This is not altogether surprising, since subjects have access to more information regarding their posttest performance.

Although analysis of the shift week as a whole indicated a moderate relationship between predicted and actual performance, results of individual shift-by-shift analysis revealed a relatively poor association between predicted and actual performance for all performance parameters except PVT. Furthermore, for all performance parameters, this association was generally higher earlier in the week, suggesting a decrease in rating over the week. A possible explanation for this is that subjects were becoming increasingly sleep deprived, thus impairing the ability to accurately rate performance. Although results indicated that there was a significant average cumulative sleep debt over the week (approximately 3.4 hours prior to the final night shift), sleep efficiency and TST remained fairly stable. As such, this seems an unlikely explanation.

An alternative reason for the decrease in rating accuracy across the week of shifts may be provided by subjects' levels of motivation. During the experimental period, subjects gave performance predictions for each performance parameter 56 times. It is possible that this resulted in decreased motivation and, therefore, decreased effort to give accurate ratings. Indeed, it is a common criticism of repeated-measures investigations that repetitive testing can lead to subject boredom and subsequent declines in motivation to perform.[41]

## CONCLUSIONS

Findings of the current study suggest that during near optimal sleeping conditions, such as those created in the laboratory, it is possible for performance on laboratory tasks to adapt and be maintained during a series of consecutive night shifts. Moreover, results highlight the potential impact of learning effects in studies involving testing periods that extend across several days. While previous studies involving periods of acute sleep deprivation have indicated that individuals can accurately predict performance decrements,[15-17] the present findings suggest that this is not the case during a simulated night-work scenario. While individuals appear to have a moderate appreciation of changes in performance over the week, their ability to accurately rate performance within individual shifts is modest at best. This finding is concerning in terms of the ability of individuals to decide when they are too impaired to work or continue to drive. Consistent with previous findings,[17] it is conceivable that performance ratings are based to a large extent on subjective feelings of alertness. However, it seems that providing workers with even a low level of performance feedback may greatly improve the accuracy of predictions for future behavior. In addition, it appears that, after testing, individuals have a more accurate perception of their performance. Therefore, since individuals can appreciate previous mistakes, they can take compensatory action to avoid future errors.

## ACKNOWLEDGMENTS

This work was supported by the Australian Research Council. We gratefully acknowledge the service provided by Compumedics, Pty., Ltd.

## REFERENCES

1. Mitler M, Carskadon AM, Czeisler CA, Dement WC, Dinges DF, Graeber RC. Catastrophes, sleep and public policy: Consensus report. Sleep 1988;11:100-9.
2. Rosekind M, Smith R, Miller D, et al. Alertness management: strategic naps in operational settings. J Sleep Res 1995;4:62-6.
3. Costa G. The impact of shift and night work on health. Appl Ergonom 1996;27:9-16.
4. Glazner LK. Shiftwork: its effect on workers. AAOHN J 1991;39:416-21.
5. Torsvall L, Akerstedt T, Gillander K, Knutsson A. Sleep on the night shift: 24-hour EEG monitoring of spontaneous sleep/wake behaviour. Psychophysiology 1989;26:352-8.

6. Kecklund G, Akerstedt T. Sleepiness in long distance truck driving: an ambulatory EEG study of night driving. Ergonomics 1993;36:1007-17.
7. Browne RC. The day and night performance of teleprinter switchboard operators. Occupat Psychol 1949;21:121-6.
8. Bjerner B, Swensson A. Shiftwork and rhythm. Acta Med Scand 1953;278:102-7.
9. Folkard S. Shiftwork and performance. In: Johnson L, Tepas D, Colquhoun W, Colligan M, eds. The Twenty-four Hour Workday: Proceedings of a Symposium on Variations in Work-Sleep Schedules. Cincinnati: Department of Health and Human Services, 1981:346-73.
10. Hildebrandt G, Rohmert W, Rutenfranz J. Twelve and twenty-four hour rhythms in error frequency of locomotive drivers and the influence of tiredness. Int J Chronobiol 1974;2:97-110.
11. Fairclough SH, Graham R. Impairment of driving performance caused by sleep deprivation or alcohol: a comparative study. Hum Factors 1999;41:118-28.
12. Harrison Y, Horne JA. The impact of sleep on decision making: a review. J Exp Psychol Appl 2000;6:236-249.
13. Brown I D. Driver fatigue. Hum Factors 1994;36:298-314.
14. Dalziel JR, Soames-Job R F. Risk-taking and fatigue in taxi drivers. In: Hartley L, ed. Managing Fatigue in Transportation. Oxford: Elsevier Science Ltd; 1998:287-97.
15. Baranski JV, Pigeau RA, Angus RG. On the ability to self-monitor cognitive performance during sleep deprivation: a calibration study. J Sleep Res 1994;3:36-44.
16. Baranski J, Pigeau R. Self-monitoring cognitive performance during sleep derivation: effects of modafinil, d-amphetamine and placebo. J Sleep Res 1997;6:84-91.
17. Dorrian J, Lamond N, Dawson D. The ability to self-monitor performance when fatigued. J Sleep Res 2000;9:137-145.
18. Akerstedt T. Work hours, sleepiness and accidents: Introduction and summary. J Sleep Res 1995;4:1-3.
19. Tepas DI, Walsh JD, Armstrong D. Comprehensive study of the sleep of shiftworkers. The twenty-four hour workday. Proceedings of a symposium on variations in work-sleep schedules. US Department of Health and Human Services 1981;419-33.
20. Tilley AJ, Wilkinson, RT, Warren, PSG, Watson, B, Drud M. The sleep and performance of shiftworkers. Hum Factors 1982;24:629-41.
21. Feyer A, Williamson A, Rassack N. The Information Processing and Performance Test Battery. Sydney: National Institute of Occupational Health and Safety; 1992.
22. Lamond N, Dawson D. Quantifying the performance impairment associated with fatigue. J Sleep Res 1999;8:255-262.
23. Dinges DF, Pack F, Williams K, et al. Cumulative sleepiness, mood disturbance and psychomotor vigilance decrements during a week of sleep restricted to 4-5 hours per night. Sleep1997;20:267-77.
24. Jewett ME, Dijk D-J, Kronauer RE, Dinges DF. Dose-response relationship between sleep duration and human psychomotor vigilance and subjective alertness. Sleep 1999;22:171-9.
25. Wilkinson RT. Interaction of lack of sleep with knowledge of results, repeated testing, and individual differences. J Exp Psychol 1961;62:263-71.
26. Rechtschaffen A, Kales A. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Los Angeles: Brain Information Service/Brain Research Institute UCLA; 1968.
27. Dinges DF, Kribbs, NB. Performing while sleepy: effects of experimentally induced sleepiness. In Monk TH, ed. Sleep, Sleepiness and Performance. Chichester: John Wiley and Sons, 1991:97-128.
28. Angus R, Heslegrave R. Effects of sleep loss on sustained cognitive performance during a command and control situation. Behav Res Meth Inst Comp 1985;17:55-67.
29. Haslam DR. Sleep loss, recovery sleep, and military performance. Ergonomics 1982;25:163-78.
30. Webb W, Levy C. Age, sleep deprivation and performance. Psychophysiology 1982; 19:272-6.
31. Horne JA. Wilkinson S. Chronic sleep reduction: daytime vigilance performance and EEG measures of sleepiness, with particular reference to “practice” effects. Psychophysiology 1985;22:69-78.
32. Kribbs NB, Dinges DF. Vigilance decrement and sleepiness. Sleep onset mechanisms. Washington: American Psychological Association; 1994:113-25.
33. Wilkinson RT. Effects of up to 60 hours sleep deprivation on different types of work. Ergonomics 1964;7:175-86.
34. Barnes RG, Deacon SJ, Forbes MJ & Arendt J. Adaptation of the 6-sulphatoxymelatonin rhythm in shiftworkers on offshore oil installations during a 2-week 12-h night shift. Neuroscience Letters 1998; 241(1): 9-12.
35. Folkard S, Monk T. Shiftwork and performance. Hum Factors 1979;21:483-92.
36. Akerstedt T. Is there an optimal sleep-wake pattern in shift work? Scand J Work Environ Health 1998;24:18-27.
37. Porcu S, Bellatreccia A, Ferrara M, Casagrande M. Sleepiness, alertness and performance during a laboratory simulation of an acute shift in the sleep/wake cycle. Ergonomics 1998;8:1192-202.
38. Broadbent DE. Performance and its measurement. Br J Clin Pharmacol 1984;18: S5-9.
39. Osberg TM, Shrauger JS. Self-prediction: exploring the parameters of accuracy. J Pers Soc Psychol 1986;51:1044-57.
40. Arnedt, JT, Wilde JS, Munt PW, Maclean AW. Simulated driving performance following prolonged wakefulness and alcohol consumption: separate and combined contributions to impairment. J Sleep Res 2000;9:233-41.
41. Babkoff H, Mikulincer M, Caspy T, Kempinski D, Sing H. The topology of performance curves during 72 hours of sleep loss: a memory and search task. Q J Exp Psychol 1988;324:737-56.