J. Sleep Res. (2000) 9, 137–144

The ability to self-monitor performance when fatigued

JILL DORRIAN, NICOLE LAMOND and DREW DAWSON
The Centre for Sleep Research, The Queen Elizabeth Hospital, Woodville, Australia

Accepted in revised form 28 November 1999; received 24 June 1999

SUMMARY The present study aimed to systematically investigate the effects of elevated fatigue levels on the ability to self-monitor performance. Eighteen participants, aged 19–26 y, remained awake for a period of 28 h. Neurobehavioural performance was measured at hourly intervals using four tests from a standardized computer test battery. From these four tests, six measures of performance were obtained: grammatical reasoning (accuracy and response latency); vigilance (accuracy and response latency); simple sensory comparison and tracking. In addition, before and after each test, participants completed visual analogue scales which required them to rate their alertness level and the speed and accuracy of their performance. Individual test results for both self-ratings and neurobehavioural performance were converted to *z*-scores. Planned comparison analysis indicated that scores on four of the six performance measures decreased significantly as hours of wakefulness increased. Similarly, predicted performance scores for all six measures of performance decreased significantly. Analysis revealed moderate correlations between predicted and actual performance for the four parameters affected by fatigue. Furthermore, moderate to high correlations were found between all six performance parameters and their respective post-test self-ratings. In addition, moderate to high correlations were found between predicted performance and alertness. Taken together, these findings suggest that as fatigue levels increase, subjects globally assess performance decrements. Results suggest that subjective alertness may in part mediate an individual's global assessment of performance.

KEYWORDS fatigue, self-monitoring, performance

INTRODUCTION

The work patterns of shift-workers frequently result in sleep of limited duration and quality (Mitler *et al.* 1988). Research indicates that this sleep loss is often associated with increased levels of fatigue and subsequently impaired neurobehavioural performance. A combination of these factors typically leads to lower productivity and higher risk of accident (Elkin and Murray 1974; Folkard and Monk 1979; Babkoff *et al.* 1985; Mitler *et al.* 1988; Rosekind *et al.* 1995). Recent studies have indicated that fatigue is often a major contributor to accident and death. For example, estimates indicate that human error is responsible for ≈65–90% of transportation accidents (Lauber and Kayten 1988). Indeed, fatigue has been identified as a major contributing factor in catastrophes such as *Exxon Valdez* and Three-Mile Island (Rosekind *et al.* 1995).

As a consequence, research has focused on methods for assessing the effects of irregular work hours and elevated fatigue on worker performance. One possible method of assessment is based on the suggestion that the ability to make accurate self-judgements may determine whether an individual is likely to engage in safe or unsafe behaviours. That is, accurate self-monitoring of behaviour may result in a reduction in risk-taking behaviours and a subsequent reduction in accident.

For example, in a study by Baranski *et al.* (1994), subjects deprived of sleep for 46 h were asked to rate their performance on a serial addition task. Overall, their findings indicated that fatigue did not effect self-monitoring of performance. Furthermore, in a later study extending these findings, subjects monitored their performance on a visual judgement task and a complex mental addition task during 64 h of sleep deprivation. Consistent with the findings of Baranski *et al.* (1994), results

Correspondence: Dr D. Dawson, The Centre for Sleep Research, The Queen Elizabeth Hospital, Woodville Rd, Woodville SA 5011, Australia. Tel.: 61 88 222 6624; fax: 61 88 222 6623; e-mail: drew.dawson@unisa.edu.au

indicated that performance self-monitoring remained reasonably unaffected by sleep deprivation. The authors concluded that in the absence of external cues, subjects had access to relatively accurate internal feedback, and thus self-ratings were a reliable indicator of overt performance (Baranski and Pigeau 1997).

In a pilot study, conducted by our research group in 1995, changes in neurobehavioural performance across 28 h of sustained wakefulness were compared with self-reported predictions of performance and subjective alertness. Performance was measured using the visual Sternberg two-choice visual and reaction time task. The findings indicated that the extent to which self-reported performance and subjective alertness predicted actual performance was equivalent. Specifically, both measures overestimated performance with less than 16 h of sustained wakefulness, and underestimated performance with greater than 16 h of wakefulness. In addition, it was found that while response latency decreased accuracy levels remained constant. This is consistent with the literature which suggests that different components of performance may differ in their sensitivity to sleep loss and fatigue, thus supporting theories that suggest that performance should not be treated as a single variable, but as a collection of measures (Broadbent 1984). Furthermore, although results of the pilot study indicated that response latency and accuracy were affected differently by increased levels of fatigue, findings suggested that subjects did not assess speed and accuracy as different performance components. In contrast, subjects rated both components of their performance identically, and consequently predictions for response latency were more accurate than those for accuracy.

Taken together, the results of this study suggest that self-ratings of performance and subjective alertness either over- or underestimate actual performance. Therefore, these measures may not be effective mechanisms for assessing the performance of fatigued individuals. However, it should be noted that only a single test was used. Therefore, it is not clear whether the findings are characteristic of all performance parameters, or whether they are limited to the particular measure chosen. Thus, the current study sought to replicate and extend the initial findings of the pilot study by systematically comparing the effects of elevated levels of fatigue on the ability to self-monitor performance on a range of neurobehavioural performance tasks.

METHOD

Subjects

Eighteen subjects, aged 19–26 y (mean, 21.17 y; SD, 2.38 y), were recruited for the study. Volunteers were required to complete a general health questionnaire and sleep/wake diary prior to the study. Subjects who had a current health problem, and/or a history of psychiatric or sleep disorders were excluded, as were subjects who smoked cigarettes or who were taking medication known to effect sleep or performance.

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. In the week prior to commencement of the experimental condition, subjects were trained individually on the performance tests in order to familiarize themselves with the tasks and to minimize improvements in performance resulting from learning. During the training session, subjects were required to repeat each test until their performance reached a plateau.

In order to produce substantial levels of fatigue, subjects were deprived of sleep for one night and performance was measured throughout the night, at the low point of the circadian cycle. On the night before the experimental session, subjects reported to the laboratory at 20.00 h. Prior to retiring at 23.00 h, subjects were required to complete additional practice trials on each tasks. Subjects were woken at 07.00 h and allowed to breakfast and shower prior to an 08.00 h baseline testing session. Subjects then completed a performance testing session every hour until the final test which commenced at 11.00 the following day (a total of 28 h of sustained wakefulness). Subjects participated in groups of four (because of last minute cancellations there were two groups of three). In between testing sessions, subjects could read, write, watch television or converse with other subjects, but were not allowed to exercise, shower or bath. Food and drinks containing caffeine were prohibited the night before and during the experimental conditions. Subjects were monitored to ensure continuous wakefulness.

Neurobehavioural tests

Neurobehavioural performance was measured using a standardized computer-based test battery (developed by Worksafe Australia, NSW, Sydney). Based on a standard information processing model (Wickens 1984), the battery sought to provide a broad sampling of various components of neurobehavioural performance. Four of 12 possible performance tests were used, such that the level of cognitive complexity ranged from simple to more complex (as listed below). As speed and accuracy scores can be affected differently by sleep loss and fatigue (Webb and Levy 1982; Angus and Heslegrave 1985), two of the four tests investigated assessed both speed and accuracy.

The simple sensory comparison task required participants to focus on an attention fixing spot displayed on the monitor for 750 ms. Following this, a line of stimulus characters, divided into three blocks of either two numbers, two letters or one of each were displayed. Participants were then required to respond to a visual cue, which appeared in the position of one of the stimulus blocks, by naming the block of stimulus characters which had previously been in that position. Verbal responses were allocated 2 points if both characters in the block were named correctly, 1 point if one of the characters was named correctly, or 0 points if neither character was named correctly. All stimulus characters were presented on the monitor for a period of 2, 3 or 5 ms, for a total of 26 trials (i.e. maximum score of 52 points).

The unpredictable tracking task (3-min trials) was performed using a joystick to control the position of a tracking cursor by centring it on a constantly moving target. Percentage of time on target was the performance measure.

The vigilance task (3.5-min trials) was performed using a response box consisting of an array of six lights, six black buttons and a red button. If a single light was illuminated, subjects were required to press the corresponding black button underneath it. If however, two lights were illuminated simultaneously, subjects were required to press the red button. Each light turned off when a response was made, or after 2500 ms. For this report, two vigilance measures were evaluated: (i) the number of correct responses (accuracy), and (ii) increases in the duration of responses (response latency).

The grammatical reasoning task was based on a similar task by Baddeley (1968). This task required subjects to decide and indicate whether a logical statement, that referred to a pair of letters, was true or false (e.g. the statement 'A precedes B' is true for the letter pair AB). For each trial, subjects were presented individually with 32 statements, beneath which were a pair of letters (either AB or BA). To respond, subjects were required to hold down a 'home' button on the response box until they were ready to press one of two other buttons, designated either 'true' or 'false'. Subjects were instructed to concentrate on accuracy, rather than speed. In this report, both accuracy (percentage of correct responses) and response latency were evaluated.

During test sessions, subjects were seated in front of the workstation 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 ≈ 15 min. Subjects received no feedback during the study, as previous research indicates that knowledge of results influences performance levels (e.g. Wilkinson 1961).

Self-rating questionnaires

Self-rating questionnaires were used to record pre- and post-test subjective ratings of alertness, response speed and response accuracy. The questionnaires consisted of seven linear, non-numeric, 100-mm visual analogue scales (VAS). Pre-test, the seven scales were administered consecutively, whereas post-test scales were administered one at a time, on a task-by-task basis. For the first question, pertaining to subjective alertness, the extreme left of the line was marked 'struggling to remain awake', and the extreme right of the line was marked 'extremely alert and wide awake.' Question 2, pertaining to the tracking task was anchored with '0% of the time on target' on the left and '100% of the time on target' on the right. Questions 4 and 6 related to accuracy on the grammatical reasoning and vigilance tasks, respectively. These lines were marked 'none correct' on the extreme left, and 'all correct' on the extreme right. Questions 5 and 7 related to response speed on grammatical reasoning and vigilance tasks, respectively. These lines were anchored with 'extremely slowly' on the left, and 'extremely quickly' on the right. Subjects were instructed to answer the pre- and post-test questions relative to their perception of their own average alertness and performance levels from the training session.

Statistical analysis

To control for interindividual variability on neurobehavioural performance, and to allow comparison of predicted and actual performance scores, all scores for each subject were converted to z-scores. Data were then collapsed into 2-h bins.

Evaluation of systematic changes in subjective alertness and each predicted and actual performance parameter across time (hours of wakefulness) were assessed separately by repeated measures analysis of variance (ANOVA), with significance levels corrected by sphericity by Greenhouse Geisser Epsilon.

Following ANOVAS, planned comparisons were conducted. For each of the dependent variables, for each performance parameter, a baseline value was calculated by averaging the values obtained at 09.00, 11.00, 13.00, 15.00 and 17.00 h. During this time, which coincides with the hours of a normal working day (i.e. '9 to 5'), scores remained relatively stable. Each of the planned comparisons focused on statistical tests comparing the baseline value with each of the remaining times (i.e. 19.00 to 11.00 h the following day). For the planned comparisons it was hypothesized that mean levels for the dependent variables (alertness ratings, actual test performance, pre-test predicted performance and post-test assessed performance), would show a significant decline over the 28-h period of sustained wakefulness.

Time series correlation coefficients were calculated for each participant –3 to + 3 time lags, for comparisons between pre-test predicted performance and pre-test subjective alertness, pre-test predicted performance and actual performance, post-test performance estimates and actual performance and post-test performance estimates and post-test alertness. 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.

RESULTS

Actual performance

Table 1 displays the results of the ANOVAS for each performance variable as a function of hours of wakefulness. Four of the six performance parameters showed statistically significant ($P = 0.0001$) variation by hours of wakefulness. Figure 1 displays the results of subsequent planned comparisons. Scores on these four performance parameters were significantly lower ($P = 0.001$) than the baseline score. In

Table 1 Summary of ANOVA results for actual performance, pre-test predictions and post-test estimates

Performance parameter	*Actual performance $F_{13,221}$*	*P*	*Pre-test predictions $F_{13,221}$*	*P*	*Post-test estimates $F_{13,221}$*	*P*
VIG accuracy	10.474	0.0001	11.477	0.0001	25.616	0.0001
VIG response latency	31.214	0.0001	23.131	0.0001	32.184	0.0001
GRG accuracy	1.688	NS	6.985	0.0001	15.538	0.0001
GRG response latency	10.007	0.0001	16.625	0.0001	16.625	0.0001
Simple sensory comparison	1.255	NS	6.413	0.0001	27.463	0.0001
Tracking	11.971	0.0001	25.195	0.0001	40.421	0.0001

VIG, vigilance; GRG, grammatical reasoning. Significance levels corrected by Greenhouse Geisser Epsilon.

general, poorest performance occurred between 05.00 and 10.00 h.

Pre-test predicted performance

The results of the ANOVAS indicated that predicted performance scores for all six performance parameters showed statistically significant ($P = 0.0001$) variation by hours of wakefulness (Table 1). As indicated in Fig. 1, predicted performance scores were significantly ($P = 0.001$) lower than baseline after 01.00, for all six of the performance parameters. In general, predicted scores were lowest between 07.00 and 10.00 h.

Correlations between predicted and actual performance

Time series correlations between predicted and actual performance scores were highest at time lag zero. Results are displayed in Table 2. Significant correlations ($P = 0.05$–0.01) were found between actual and predicted performance for three of the six performance parameters, with a fourth approaching significance (i.e. correlations for the tracking task and the response latency components of the vigilance and grammatical reasoning tasks were significant, and the correlation for the accuracy component of the vigilance task was approaching significance $r = 0.47$, critical $r_{(12)} = 0.53$).

Pre- and post-test subjective alertness

Analysis of variance revealed statistically significant variation in pre-test ($F_{9,153} = 52.362$, $P = 0.0001$) and post-test ($F_{9,153} = 45.751$, $P = 0.0001$) subjective alertness by hours of wakefulness. As illustrated in Fig. 2, planned comparison analysis indicated that pre- and post-test subjective alertness scores were significantly lower than baseline from 01.00 onwards. It was found that subjective alertness reached a nadir at 07.00 h.

Time series correlation analysis between pre- and post-test ratings of subjective alertness were highest at time lag zero. A significant correlation ($r = 0.75$, $P < 0.01$) was found between the two measures.

The results of an ANOVA indicated no statistically significant difference ($F_{1,34} = 0.64$, $P = 0.8021$) between pre- and post-test subjective alertness ratings.

Correlations between pre-test predicted performance and subjective alertness

Time series correlations between predicted performance and subjective alertness scores were highest at time lag zero. Results are displayed in Table 3. Significant correlations ($P = 0.05{-}0.01$) were found between predicted performance and subjective alertness for all six performance parameters.

Post-test performance self-ratings

The results of the ANOVAS indicated that post-test performance estimates for all six performance parameters showed statistically significant ($P = 0.0001$) variation by hours of wakefulness (Table 1). Planned comparison analysis indicated that post-test self-ratings for all of the performance parameters were significantly lower ($P = 0.001$) than baseline (Fig. 1). In general, predicted scores were lowest between 03.00 and 10.00 h.

Time series correlations between post-test self-ratings and actual performance were highest at time lag zero. As indicated in Table 2, analysis revealed significant correlations ($P = 0.05{-}0.01$) between post-test self-ratings and actual performance for five of the six performance parameters.

Time series correlations between post-test performance ratings and post-test subjective alertness ratings were highest at time lag zero. As indicated in Table 4, analysis revealed significant correlations ($P = 0.05{-}0.01$) between post-test self-ratings and of performance and alertness for five of the six performance parameters.

DISCUSSION

As expected, in this study increased levels of fatigue had a clearly measurable effect on neurobehavioural performance. We observed that as hours of wakefulness increased, performance levels for four of the six parameters decreased significantly. Similarly, as hours of wakefulness increased, subjective ratings of performance and alertness decreased. Comparison of actual and predicted performance indicated moderate correlations for the four performance parameters that decreased significantly with increasing fatigue. In addition, for all six parameters moderate to high correlations were found between performance self-ratings and subjective alertness.

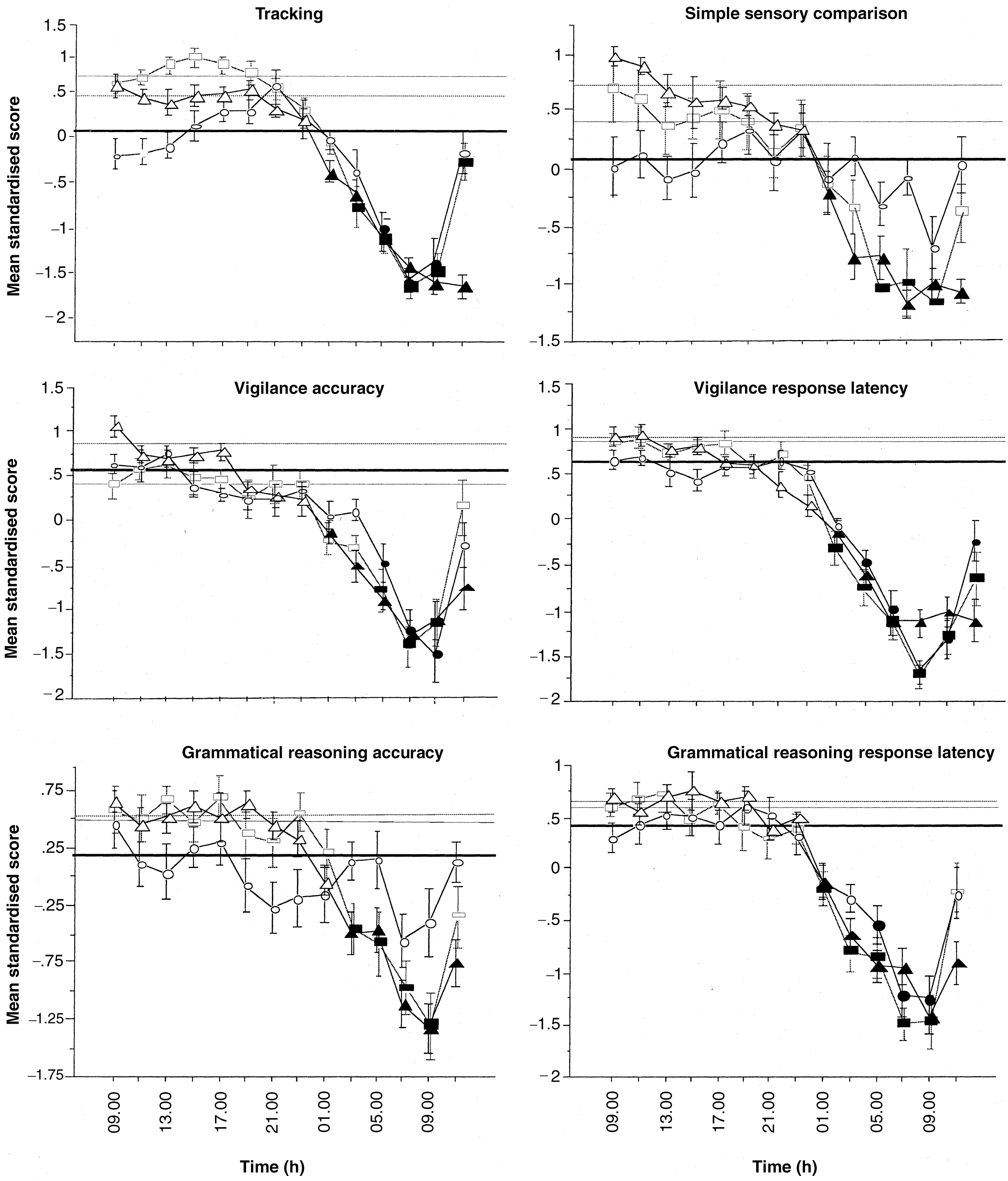

Figure 1. Actual performance (circles), pre-test predicted performance (triangles) and post-test performance estimates (rectangles) for each parameter. The solid line represents the baseline value for actual performance; the dotted line represents the baseline value for pre-test predicted performance; the hairline represents the baseline value for post-test performance estimates. The filled circles, triangles and rectangles indicate time points at which actual performance and pre- and post-test performance estimates significantly ($P < 0.001$) differ from the baseline.

Results of our pilot study (1995), suggested that self-ratings of performance and subjective alertness were not accurate indicators of overt performance. However, the findings of this study suggest that this may not be the case. We observed that for the parameters that were affected by fatigue, predicted performance closely tracked actual performance. This suggests

Table 2 Results of time series correlations between actual performance scores and both pre-test predicted and post-test estimated performance scores

Actual performance	*Pre-test predictions*	*Post-test estimates*
VIG accuracy	0.47	0.65*
VIG response latency	0.72**	0.81**
GRG accuracy	0.22	0.54*
GRG response latency	0.60*	0.78**
Simple sensory comparison	0.17	0.42
Tracking	0.61*	0.67**

Mean r-values across all participants are shown for each test at time lag zero. $^*P < 0.05$, $^{**}P < 0.01$. VIG, vigilance; GRG, grammatical reasoning.

Table 3 Results of time series correlations between pre-test performance predictions and subjective alertness ratings

Pre-test predicted performance	*Pre-test subjective alertness*
VIG accuracy	0.81**
VIG response latency	0.85**
GRG accuracy	0.58*
GRG response latency	0.69**
Simple sensory comparison	0.85**
Tracking	0.90**

Mean r-values across all participants are shown for each test at time lag zero. $^*P < 0.05$, $^{**}P < 0.01$. VIG, vigilance; GRG, grammatical reasoning.

that, consistent with the findings of Baranski *et al.* (1994), subjects were able to accurately monitor the performance impairment associated with increased levels of fatigue.

Interestingly, this was not observed for the two performance parameters that were not significantly affected by fatigue. Given that two of the performance parameters remained relatively unaffected, it was expected that this would be reflected in the respective pre-test self-ratings. However, contrary to this expectation, participants gave equivalent performance self-ratings for each parameter. As such, performance on these two parameters was not predicted accurately. Hence, it is apparent that rather than making test-specific predictions, individuals were globally assessing future performance. However, it is important to note that pre-test performance estimates were given together at the start of each test battery rather than before each individual task. Thus, equivalent or global estimates may be the result of the way in which the scales were administered. To control for this in future studies it is important to administer pre-test rating scales before the particular performance task to which they apply.

Consistent with previous studies (e.g. Gillberg *et al.* 1994), we observed that subjective alertness (measured pre- and post-test) decreased significantly as hours of wakefulness increased. Furthermore, this decrease in subjective alertness coincided with the observed decrease in pre- and post-test performance self-ratings, such that subjective ratings of performance and alertness tracked each other closely. Although correlation analysis can in no way determine direction of causality, we think it possible that pre- and post-test performance ratings

Table 4 Results of time series correlations between post-test performance estimates and post-test subjective alertness ratings

Post-test predicted performance	*Post-test subjective alertness*
VIG accuracy	0.42
VIG response latency	0.86**
GRG accuracy	0.59*
GRG response latency	0.77**
Simple sensory comparison	0.56*
Tracking	0.85**

Mean r-values across all participants are shown for each test at time lag zero. $^*P < 0.05$, $^{**}P < 0.01$. VIG, vigilance; GRG, grammatical reasoning.

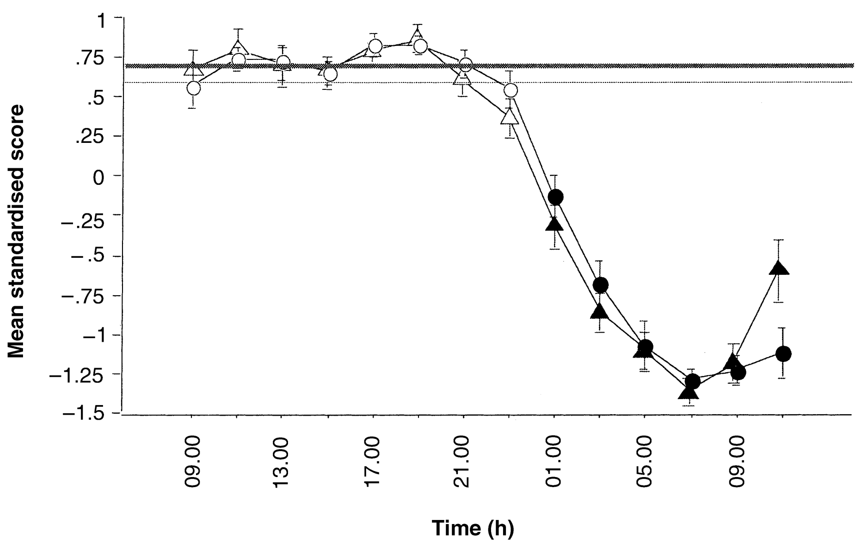

Figure 2. Pre-test (circles) and post-test (triangles) subjective ratings of alertness. The hairline represents the baseline value for pre-test alertness and the solid line represents the baseline for post-test alertness. The filled circles and triangles indicate time points at which subjective alertness ratings significantly differ ($P < 0.001$) from the baseline.

were affected to some degree by alertness levels. Indeed, subjective alertness is likely to be a reliable cue for performance judgement. That is, as their fatigue levels increased, subjects typically expected that their performance would decrease. This suggests that subjective alertness may, in part, mediate an individual's assessment of performance.

In addition, the results suggested a strong relationship between pre- and post-test subjective alertness ratings. Indeed, no statistically significant differences were found between the two sets of scores. This suggests first, that there was no significant 'break effect' (i.e. elevated alertness at the beginning of each testing session resulting from the preceding 45-min break in testing), and secondly, that the testing sessions themselves did not significantly affect alertness levels.

Several consistencies were found between this study and our earlier pilot study. First, a strong relationship was found between subjective ratings of alertness and predicted performance in both studies, suggesting that individuals expect actual performance to be linked with alertness levels. Secondly, while the specific components of performance differed in their degree of sensitivity, in general, fatigue had a detrimental effect on neurobehavioural performance.

Specifically, the results of the pilot study indicated a differential effect of increased fatigue on response latency and accuracy. That is, while accuracy levels remained relatively constant, response latency declined with increasing fatigue. Similarly, in the current study, scores on the accuracy component of the grammatical reasoning task remained relatively unaffected, whereas scores on the response latency component declined significantly during the 28-h period. The most likely explanation for this is that subjects were instructed to focus on accuracy at the expense of speed on all tests. Alternatively, this result may reflect a natural speed/accuracy trade off, which is often observed for similar self-paced tasks (Haslam 1982; Webb and Levy 1982; Angus and Heslegrave 1985).

Interestingly, we observed that performance on the simple sensory comparison task also remained relatively unaffected by increased fatigue. Of the four tests used, simple sensory comparison was the least complex, and the least monotonous. Thus, these results are in line with the suggestion that simple tasks are less sensitive to sleep deprivation (Johnson 1982). Indeed, we believe it likely that impairment of performance on this task may have occurred if we had extended the period of sustained wakefulness. It is interesting to note that several studies (e.g. Dinges *et al.* 1988) have reported that tasks similarly lacking in complexity, such as simple reaction time tasks, are affected early and profoundly by sleep loss, thus strongly suggesting that monotony may increase sensitivity to sustained wakefulness. Indeed, the fact that this task was not vulnerable to fatigue may possibly be explained by the interesting and challenging properties of the task. However, it is important to note that during the experimental period, each test was administered 28 times. Previous research suggests as the number of trials increases, the novelty of the test will decrease, and subsequently motivation will decrease. Thus, the effect of sleep deprivation on test performance increases (Wilkinson 1961; Horne and Petitt 1985).

Indeed, motivation has a profound effect on task performance. Research indicates that motivated subjects are able to increase effort in order to compensate for the effects of sleep deprivation (Dinges and Kribbs 1991). An example of such an effect was observed during this study. A clear increase in actual performance scores occurred during the 10.00 to 11.00 h trials. It is probable that this 'end of test burst' reflects the point at which subjects became aware that the study was almost over and therefore felt motivated, such that their task performance improved. Interestingly, this increase is not evident in the pre-test performance ratings.

While several studies have investigated pre-test predictions, only one (Baranski and Pigeau 1997) investigated the accuracy of post-test self-ratings. As awareness of performance impairment associated with sleep loss and fatigue may result in compensatory efforts to avoid future errors, assessing the extent of this awareness is important. Results of the current study indicate that post-test self-ratings were better predictors of actual performance than were pre-test self-ratings. This is highlighted by the 'end of test burst' which was not reflected by the pre-test self-ratings, but which was appreciated by subjects post-test. This suggests an ability to judge performance directly. Given that participants had access to more information regarding their performance when making post-test assessments, this is not altogether surprising. Indeed, while pre- and post-test judgements appear to be mediated to some extent by subjective alertness, it is probable that post-test judgements were based to a greater degree on actual performance.

The findings of the current study are encouraging when considered in terms of risk management in the workplace. Taken together, the findings suggest that, in many cases, worker self-monitoring may provide an accurate indication of the level of worker performance. Therefore, self-monitoring of performance may serve as an effective mechanism for maintaining productivity, and reducing the risk of workplace accident. Nevertheless, because not all measures of performance are affected by increased levels of fatigue in the same way, the global rating given by the individual may at times overestimate performance impairment.

ACKNOWLEDGEMENT

This research was supported by the Australian Research Council.

REFERENCES

Angus, R. and Heslegrave, R. Effects of sleep loss on sustained cognitive performance during a command and control situation. *Behav Res. Methods Instrum. Comput.*, 1985, 17: 55–67.

Babkoff, H., Genser, S. G., Sing, H., Thorne, D. R. and Hegge, F. W. The effects of progressive sleep loss on a lexical decision task: responses, lapses and response accuracy. *Behav Res. Methods Instrum. Comput.*, 1985, 17: 614–622.

Baddeley, A. D. A 3 minute reasoning task based on grammatical transformation. *Psychonomic Sci.*, 1968, 10: 341–342.

Baranski, J. and Pigeau, R. Self-monitoring cognitive performance during sleep derivation: effects of Modafinil, D-amphetamine and placebo. *J. Sleep Res.*, 1997, 6: 84–91.

Baranski, J. V., Pigeau, R. A. and Angus, R. G. On the ability to self-monitor cognitive performance during sleep deprivation: a calibration study. *J. Sleep Res.*, 1994, 3: 36–44.

Broadbent, D. E. Performance and its measurement. *Br. J. Clin. Pharmacol.*, 1984, 18: 5S–9S.

Dinges, D. F. and Kribbs, N. B. Performing while sleepy: effects of experimentally-induced sleepiness. In: Monk, T. H. (Ed.) *Sleep, Sleepiness and Performance*. John Wiley and Sons Ltd, New York, 1991.

Dinges, D. F., Whitehouse, W. G., Orne, E. C. and Orne, M. T. The benefits of a nap during prolonged work and wakefulness. *Work Stress*, 1988, 2: 139–153.

Elkin, A. J. and Murray, D. J. The effects of sleep loss on short-term recognition memory. *Can J. Psychology*, 1974, 28: 192–198.

Folkard, S. and Monk, T. Shiftwork and performance. *Hum. Factors*, 1979, 21: 483–492.

Gillberg, M., Kecklund, G. and Akerstedt, T. Relations between performance and subjective ratings of sleepiness during a night awake. *Sleep*, 1994, 17: 236–241.

Haslam, D. R. Sleep loss, recovery sleep and military performance. *Ergonomics*, 1982, 25: 163–178.

Horne, J. A. and Petitt, A. N. High incentive effects on vigilance performance during 72 hours of total sleep deprivation. *Acta Psychologica*, 1985, 58: 123–129.

Johnson, L. C. Sleep deprivation and performance. In: W. B. Webb (Ed.) *Biological Rhythms, Sleep and Performance*. Wiley, New York, 1982.

Lauber, J. K. and Kayten, P. J. Sleepiness, circadian dysrhythmia, and fatigue in transportation accidents. *Sleep*, 1988, 11: 503–512.

Mitler, M., Carskadon, A. M., Czeisler, C. A., Dement, W. C., Dinges, D. F. and Graeber, R. C. Catastrophes, sleep and public policy: consensus report. *Sleep*, 1988, 11: 100–109.

Rosekind, M., Smith, R., Miller, D., Co, E., Gregory, K., Webbon, L., Gander, P. and Lebacq, Z. Alertness management: strategic naps in operational settings. *J. Sleep Res.*, 1995, 4 (Suppl.): 62–66.

Webb, W. and Levy, C. Age, sleep deprivation and performance. *Psychophysiology*, 1982, 19: 272–276.

Wickens, C. D. *Engineering Psychology and Human Performance*. Scott Foresman, Boston, 1984.

Wilkinson, R. T. Interaction of lack of sleep with knowledge of results, repeated testing and individual differences. *J. Exp Psychol.*, 1961, 62: 263–271.