Applied Ergonomics 38 (2007) 155–166

APPLIED ERGONOMICS

www.elsevier.com/locate/apergo

Simulated train driving: Fatigue, self-awareness and cognitive disengagement

Jillian Dorrian*, Gregory D. Roach, Adam Fletcher, Drew Dawson

The School of Psychology and the Centre for Sleep Research, The University of South Australia, 7th Floor Playford Building (P7-35), City East Campus, Frome Rd, Adelaide, SA 5000, Australia

Received 22 December 2004; accepted 17 March 2006

Abstract

Fatigue is a serious issue for the rail industry, increasing inefficiency and accident risk. The performance of 20 train drivers in a rail simulator was investigated at low, moderate and high fatigue levels. Psychomotor vigilance (PVT), self-rated performance and subjective alertness were also assessed. Alertness, PVT reaction times, extreme speed violations (>25% above the limit) and penalty brake applications increased with increasing fatigue level. In contrast, fuel use, draft (stretch) forces and braking errors were highest at moderate fatigue levels. Thus, at high fatigue levels, errors involving a failure to act (errors of omission) increased, whereas incorrect responses (errors of commission) decreased. The differential effect of fatigue on error types can be explained through a *cognitive disengagement* with the virtual train at high fatigue levels. Interaction with the train reduced dramatically, and accident risk increased. Awareness of fatigue-related performance changes was moderate at best. These findings are of operational concern.

Keywords: Performance; Fatigue; Self-awareness

1. Introduction

1.1. Train drivers, sleep loss and fatigue

Fatigue is an important issue for the rail industry, with train drivers' schedules resulting in sleep-related problems (Foret and Latin, 1972; Pilcher and Coplen, 2000; Roach et al., 2003). Research has identified several factors responsible for elevated fatigue among train drivers, including uncertain shift times, long commutes, suboptimal terminal sleeping conditions, and the fact that some train drivers may not have daytime rest before a night shift (Pollard, 1991). Indeed, Pollard (1996) found that for shifts that started between 10pm and 4am, train drivers reported that they slept fewer than 6 h per day.

Simulator studies suggest that driving a car, truck or aeroplane is impaired by sleep loss and fatigue (Arnedt et al., 2000; Caldwell et al., 2000; Fairclough and Graham, 1999; Gillberg et al., 1994). While there are a paucity of journal articles on fatigue and performance in rail simulators, industry reports detail testing in this area. For example, Thomas and Raslear (DOT/FRA, 1997) investigated the effects of backward rotating work schedules on the fatigue and performance of 55 train drivers during a simulated 190-mile run. These schedules resulted in a cumulative sleep debt, which was associated with decreased alertness, increased fuel use, missed alerter signals and failures to sound the horn at grade crossings.

Further, investigations have identified sleep loss and fatigue as contributing factors in numerous rail accidents. For example, the 1997 coal train collision in Beresford, Australia, seriously injured three people, and caused extensive damage to the trains, station platform and associated structures. The investigation found that work-related fatigue was a major contributor to the accident (Pearce, 1999). Similar incidents in 1984, in Wiggins, Colorado, and in Newcastle, Wyoming resulted in a total of four injuries and seven deaths with a combined economic cost estimated at over $US5 million (Lauber and Kayten, 1988). Railway accident analyses in Japan (Kogi and Ohta, 1975) and China (Zhou, 1991) revealed

*Corresponding author. Tel.: +61 8 8302 6624; fax: +61 8 8302 6623.
E-mail address: jill.dorrian@unisa.edu.au (J. Dorrian).

0003-6870/$ - see front matter
doi:10.1016/j.apergo.2006.03.006

that fatigue and sleepiness were among the causes. Thus, sleep loss and fatigue are becoming recognised internationally as fundamental safety problems in the rail industry, associated with serious social and economic cost (Edkins and Pollock, 1997).

1.2. Theories underlying the neurobehavioural effects of sleep loss and fatigue

Since the earliest studies of sleep loss (Kleitman, 1963; Patrick and Gilbert, 1896), periods of non-responding, or performance 'lapses' have been observed. These findings lead to the 'lapse hypothesis' to explain the effects of sleep loss on performance (Williams et al., 1959). This approach contends that sleep loss produces performance lapses, which occur more frequently with increasing time awake. However, between lapses individuals may perform near optimally, rendering performance continually more variable. While the occurrence of lapses has been well-documented (Dinges and Powell, 1988; Doran et al., 2001; Kleitman, 1963; Patrick and Gilbert, 1896), the lapse hypothesis has several shortcomings (reviewed in Dinges, 1992). Specifically, it fails to account for several phenomena including slowing of the fastest 10–20% of reaction times (Dinges and Powell, 1989), performance deterioration with time-on-task (Dinges and Powell, 1988) and differential impairment observed for tasks with varying characteristics (Kjellberg, 1977). In addition, while the lapse hypothesis centres around failures to respond, studies have indicated that both errors of omission (lapses) and errors of commission (responses in the absence of a stimulus) increase with increasing hours of wakefulness (Doran et al., 2001).

To account for this, an alternative hypothesis has been developed which focuses on the concomitant increase in errors of commission, arguing that it reflects compensatory effort to counteract the effects of sleep loss (Doran et al., 2001). That is, during situations of sleep loss an individual may be overcome by sleepiness, up to the point where they fall asleep. Alternatively, they may engage in compensatory behaviour, such that their performance will be maintained, at least temporarily. This is referred to as 'state instability' which is considered to be responsible for increasing performance variability during conditions of sleep deprivation.

Both the lapse hypothesis and the state instability hypothesis have been largely generated from research using laboratory tasks, measuring basic aspects of performance such as reaction time and sustained attention (Doran et al., 2001; Kleitman, 1963; Patrick and Gilbert, 1896). These assessments are valuable as they focus on underlying aspects of performance that are fundamental to any task. While studies using such tasks have shaped our understanding of the performance impairment experienced by a fatigued individual, they have also been criticised for providing uni-dimensional, artificial assessments of performance. Since real-world performance is multi-dimensional, the way in which such theories will apply to more operational performance assessment is unclear. For example, train driving is a multifaceted task, which relies on numerous aspects of neurobehavioural functioning including sustained attention, memory and planning (Roth, 2000). Given the complexities of the train environment, the relationship between sleep loss, errors of omission and errors of commission may not represent a direct reflection of laboratory findings.

In particular, potential mechanisms for triggering compensatory effort in this environment may be questioned. Previous work suggests that a fatigued individual will maintain a high level of performance on tasks that are most critical to safety at the expense of others (Fairclough and Graham, 1999; Hockey, 1997). Importantly, it has been suggested that this safety-protecting, compensatory response to increasing fatigue is cued by self-awareness of rising performance impairment (Fairclough and Graham, 1999). Certainly, laboratory studies have demonstrated that the ability to self-monitor performance impairment on computer-administered tasks remains intact during periods of acute sleep loss (Baranski et al., 1994; Baranski and Pigeau, 1997; Dorrian et al., 2000). However, a recent laboratory study involving simulated nightshifts found only a moderate relationship between self-ratings of performance and actual performance on a battery of tasks (Dorrian et al., 2003). Further, results of car simulator studies in this area have been inconsistent. While Fairclough and Graham (1999) concluded that participants were aware of reduced performance efficiency when fatigued, Arnedt and colleagues (2000) observed little relationship between simultaneous performance self-ratings and driving behaviour. Thus, it is not clear whether fatigued train drivers would have accurate performance insight, on which to base a compensatory response to rising fatigue.

Given the association between sleep loss, fatigue, performance impairment and accidents in rail, this is an important area for future research. Thus, the broad aim of this study was to investigate the effects of sleep loss and fatigue on performance in a rail simulator. Specifically, the research aimed to address the way in which sleep loss and fatigue affect: (1) errors of omission and commission in the virtual train; and (2) the ability to self-monitor train driving performance.

2. Methods

2.1. Subjects

Twenty male train drivers, recruited from four Queensland depots (Rockhampton, Emerald, Gladstone and Bluff) participated in the study. Each subject completed a general health questionnaire before commencing the study, and a Sleep/Wake Diary throughout the data collection period.

2.2. Procedure

Testing was conducted at the Queensland Rail Driver Training Centre (DTC), Rockhampton. Ethics approval was granted by the North Western Allied Health Ethics of Human Research Committee and The University of South Australia Human Research Ethics Committee.

Participants attended the DTC on three occasions, as indicated in Fig. 1. The first was a training session, designed to minimise practise effects. During training, participants observed one trip over the selected track section in the rail simulator, driven by a driver trainer (approximately 100 min in length). The trainer explained their driving strategy, and the participants were instructed to comply with that strategy. Participants then drove at least two further trips themselves. They were encouraged to make track notes, as they would when learning any new piece of track, for reference during the experimental conditions. Participants also completed at least three trials on a 10-min Psychomotor Vigilance Task (PVT, Dinges and Powell, 1985), as research indicates that the PVT has a 1–3 trial learning curve (Dorrian et al., 2004).

Participants then returned to the DTC for two testing periods, designed to produce varying levels of fatigue. Each testing period consisted of an 8-h 'shift', which was incorporated into the participants' normal work schedule. A daytime testing period was conducted between 1000 and 1800 h (at the high point in the circadian cycle) following an adequate night's sleep. A nighttime testing period was conducted between 2300 and 0700 h, after drivers had worked at least two consecutive night shifts. Each period consisted of four 2-h sessions that included the 100-min trip in the rail simulator and the 10-min PVT (see Fig. 1). In accordance with award regulations, there was a 20-min meal break between sessions one and two, and a 10-min rest break between sessions three and four.

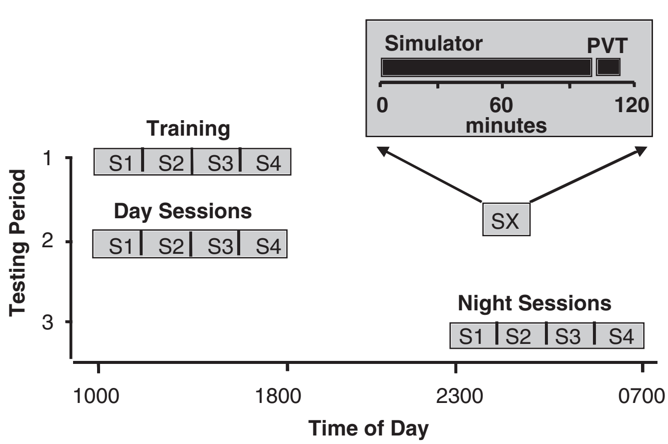

Fig. 1. Protocol diagram indicating temporal placement of training and day and night testing periods. Each testing period consists of 4, 2 h sessions (S1–S4). The top right-hand portion of the figure illustrates the composition of each 2 h session (SX).

2.3. Rail simulator

Two rail simulators (housed in separate rooms) consisted of a realistic cabin with fully operational control panels and authentic sound. Cabins faced a 3×3 m screen, onto which track footage was projected. Progress was monitored from an adjacent control room. The virtual train was a freight 2100 class diesel locomotive, measuring 432 m, weighing 1097 tonnes, with a single locomotive hauling 25 wagons. This type of train was chosen due the relative difficulty associated with manoeuvring a short, heavy train. Participants drove a 62 km piece of track between Littabella and Netley Stations on the Bundaberg–Gladstone line (approximately 100-min of driving). The track contained 11 stations, 22 bridges, and 6 hills. Speed was restricted to 70 km/h in 2 areas, to 60 km/h in 14 areas, to 50 km/h in 6 areas and to 40 km/h in 8 areas, with a maximum track speed of 80 km/h for this type of train. This track section was selected to provide areas where a high level of train manipulation was required, and areas requiring a low level of interaction with the virtual train.

General driving measures included:

- fuel use (l);
- trip time (s); and
- buff and draft forces (kN) [i.e. forces exerted as wagons bunch together and stretch apart, respectively].

Errors in brake and throttle manipulation included:

- Heavy brake reductions [i.e. the brakes are applied too strongly (>70 kPa) while the train speed is too great (>20 km/h)].
- Failures to make a split service (*no split service*) [i.e. the brakes are reapplied too early following an initial brake application (<1.5 s later)].
- Over utilisation of the brake [i.e. depleting the brake pipe pressure such that no further brake applications are possible (<360 kPa)].
- Fast throttle changes [i.e. throttle position changes within 3 s of each other].

Driving violations included:

- failures to respond to the in-cab station protection and vigilance systems, resulting in penalty (automatic emergency) brake applications (i.e. bringing the virtual train to a complete stop);
- minor speed violations [$<10\%$ above the speed limit];
- moderate speed violations [10–25% above the limit]; and
- extreme speed violations [$>25\%$ above the limit].

2.4. Self-rating scales

Before and after each PVT and driving session, participants completed a questionnaire consisting of 100 mm Visual Analogue Scales (VAS) anchored with semantic opposites. Participants were instructed to complete the ratings relative to their perception of their average performance levels from the training session. Participants gave an overall rating of driving performance, as well as specific ratings of brake, throttle and fuel efficiency, buff and draft forces, and trip time. Inefficient or improper brake use was calculated by adding the number of heavy brake reductions, the number of *no split service* reductions and the number of times the brake was over utilised. The number of fast throttle changes was used as a measure of inefficient or improper throttle use. Participants also rated their track rule compliance—calculated as the number of penalty brake applications plus the number of speeding violations. Speed violations were weighted by severity such that the resulting formula was, 'number of penalty brakes + number of minor speed violations + 2(number of moderate speed violations) + 3(number of extreme speed violations).' For directional equivalence between performance ratings and actual performance, measures of brake, fuel and throttle efficiency and rule compliance were multiplied by -1.

2.5. Fatigue level

During each session, subjects rated their alertness level using 100 mm VAS, anchored with 'struggling to remain awake' on the left, and 'extremely alert and wide awake' on the right. In addition, subjects average fatigue score for each session was calculated with Dawson and Fletcher's (2001) fatigue model. The basis of this model is the principle that fatigue can be produced by work, and reduced by rest or recovery. Importantly, the magnitude of fatigue production and recovery is dependant on the duration, circadian timing and recency of the work and rest periods. Using subjects' 7-day work history, the model assigns values to work and recovery (non-work) periods and, using a fatigue algorithm, calculates an expected fatigue output (for further model information, the reader is directed to Dawson and Fletcher, 2001; Fletcher and Dawson, 2001; Roach et al., 2004).

2.6. Statistical analysis

To assess the effects of fatigue on performance, each session (100 min of driving plus 10 min PVT) was assigned to one of three groups according to the average of the two corresponding hourly fatigue scores: low, moderate and high. Low fatigue encompassed fatigue scores from 0 to 40 (average $\pm$ SD: fatigue score $= 25.3 \pm 8.7$, subjective alertness $= 60.0 \pm 19.5$). Moderate fatigue included scores from 40 to 80 (fatigue score $= 60.7 \pm 11.7$, subjective alertness $= 47.0 \pm 20.2$). High fatigue was designated as any score greater than 80 (fatigue score $= 93.5 \pm 11.0$, subjective alertness $= 30.0 \pm 19.9$). As previously described (Fletcher and Dawson, 2001), 'high fatigue' was determined using data from Dawson and Reid (1997), who compared cognitive performance impairment during 28-h of sleep deprivation with impairment resulting from alcohol intoxication. A fatigue score of 80 was produced after 21–22 h of wakefulness, with performance impairment equivalent to that produced at a blood alcohol concentration of greater than 0.05% (the legal driving limit in Australia). As a relative comparison, low fatigue scores (0–40) are produced by the model for a 'standard' 0900–1700 h, Monday–Friday roster.

As outlined above, the fatigue group cut-off values were chosen for conceptual reasons. This method of splitting the data resulted in groups of uneven size (low fatigue, $n = 72$; moderate fatigue, $n = 50$; high fatigue, $n = 38$), and thus potential arises for violation of the assumption of homogeneity of variance (and therefore inflated risk of a Type 1 error). Group numbers could have been equalized by randomly excluding cases in low and moderate groups. However, the following checks were conducted to ensure robustness: (1) the ratio of largest to smallest group size was less than 4:1, and (2) the ratio of variances in each cell was less than 10:1 (Tabachnick and Fidell, 1996). Since these conditions were met, cases were not excluded to avoid unnecessary data loss.

Importantly however, this method of splitting the data also resulted in repeated measurements for individual drivers within each fatigue group. To account for this, linear mixed effects models were used, controlling for intercorrelated observations within drivers across the track. Post-hoc contrasts were specified between levels of the fixed effect factor (fatigue level). Uncorrected degrees of freedom are reported. In order to examine the changes in alertness, PVT score, simulator measures and performance rating scales, mixed models were conducted with these parameters specified as dependant variables, fatigue group as a fixed effect and subjects as a random effect. According to standard methodology, PVT data were transformed to 1/RT to correct for proportionality between the mean and SD (Dorrian et al., 2004). In order to examine relationships between performance ratings, actual performance and subjective alertness, mixed models were conducted with performance scores/alertness ratings as dependant variables, pre-/post-test performance ratings as fixed effects and subjects as a random effect. To compare these models, Akaike's Information Criterion (AIC) values are also reported, with smaller values indicating better model fit. Pearson r correlation coefficients were calculated to illustrate the degree of these relationships. However, it is important to note that these values may over or underestimate relationships as they do not control for repeated measurements by the same subject.

3. Results

3.1. Fatigue score, alertness and PVT

Fatigue scores for each group significantly differed ($p < 0.01$). Subjective alertness and PVT performance significantly decreased ($p<0.01$) with increasing fatigue level (Table 1, Fig. 2).

3.2. General simulator measures

Fuel use significantly varied ($p<0.01$) with fatigue level. A slight (non-significant) increase was observed from low to moderate fatigue levels, with a subsequent significant decrease ($p<0.05$) from moderate to high fatigue levels. Variation in trip time was significant ($p<0.01$), with longest times recorded at moderate fatigue. Peak buff and draft forces were smallest at a moderate fatigue level. This variation was significant ($p<0.05$) for draft forces only (Table 1, Fig. 2).

3.3. Errors in brake and throttle manipulation

The number of times that a heavy reduction, a *no split service* reduction, and brake over utilisation occurred was highest at a moderate fatigue level (Fig. 3). This variation was statistically significant ($p<0.01$) for the number of times the brake pressure was over utilised ($p<0.05$), with post-hoc comparisons showing that at moderate and high levels of fatigue, the incidence was significantly higher than at low levels ($p<0.05$). The number of fast throttle changes did not significantly vary with fatigue level (Table 1).

3.4. Penalty brake applications and speed violations

The number of penalty brake applications significantly ($p<0.05$) increased with increasing fatigue level (Table 1). Post-hoc statistics indicated a significantly ($p<0.05$) greater number occurred at moderate and high than at low fatigue levels (Fig. 3). Variation in the number of speed violations $<10\%$ above the limit was significant ($p<0.05$), with those at a moderate fatigue level significantly lower than those at a low fatigue level ($p<0.05$). The number of moderate speed violations increased across fatigue levels. This was approaching significance ($p=0.089$). The number of extreme violations was significantly greater ($p<0.01$) at moderate and high levels of fatigue than at low levels (Table 1, Fig. 3).

Table 1
Mixed model results for fatigue score, alertness, PVT and simulator driving measures [random effect: subjects, fixed effect: fatigue group, dependent variables: column 2]

		$F_{(2,157)}$	P
Fatigue level	Fatigue score	579.94	<0.01
	Subjective alertness	32.31	<0.01
PVT	1/RT	67.03	<0.01
General driving measures	Fuel	3.15	<0.05
	Trip time	5.20	<0.01
	Peak buff	0.89	NS
	Peak draft	4.53	<0.05
Errors and violations	Heavy reduction	1.69	NS
	No split service	0.52	NS
	BPP under	4.40	<0.05
	Fast throttle	1.24	NS
	Penalty brake	4.39	<0.05
	Speeding <10%	3.12	<0.05
	Speeding 10–25%	2.46	0.09
	Speeding >25%	8.76	<0.01

3.5. Self-ratings of performance

Pre- and post-test ratings for all measures tracked each other very closely (Fig. 4). Ratings of PVT performance significantly decreased ($p<0.01$) with increasing fatigue level. Taken together, ratings for overall performance, brake efficiency, throttle efficiency, rule violations and trip time showed significant variation ($p<0.05$) with fatigue level, such that scores indicated decreased performance on these parameters from low to moderate levels of fatigue, and equivalent, or slightly increased performance from moderate to high fatigue levels. Variation in ratings for fuel efficiency and rule violations was not significant (Table 2).

Mixed model analysis indicated that pre- and post-test PVT ratings and post-test brake and rule violation ratings were significant predictors of actual performance ($p<0.01$, Table 3). Examination of AIC for each model indicates that in general, post-test ratings contributed to models with better fit than pre-test ratings. All pre- and post-test ratings were significant predictors of subjective alertness ratings ($p<0.05$). AIC criteria indicated that models of alertness with post-test ratings as predictors were of roughly equivalent, or worse fit than pre-test rating models.

In general, Pearson r correlations reflected results of the mixed model analyses. There was a stronger relationship between predicted and actual performance for PVT than for driving measures. Apart from PVT and brake use, post-test ratings were more highly related to actual performance than pre-test ratings. With the exception of trip time, the relationship between post-test ratings and alertness was lower than that for pre-test ratings and alertness (Table 4).

4. Discussion

4.1. Performance impairment and patterns in error-making

As demonstrated previously, increasing fatigue resulted in impaired alertness (Dorrian et al., 2000; Gillberg et al., 1994) and sustained attention (Dinges and Powell, 1988, 1989; Doran et al., 2001). From such previous research, we expected to observe similar, clear declines in simulator driving performance. That is, we would have expected that general measures of driving performance would show reduced efficiency with increasing fatigue. However, this was not the case. As predicted, fuel use and trip time increased from a low to moderate fatigue level. However, from moderate to high levels of fatigue, fuel use and trip time decreased. In addition, the number of braking errors increased from low to moderate fatigue levels and

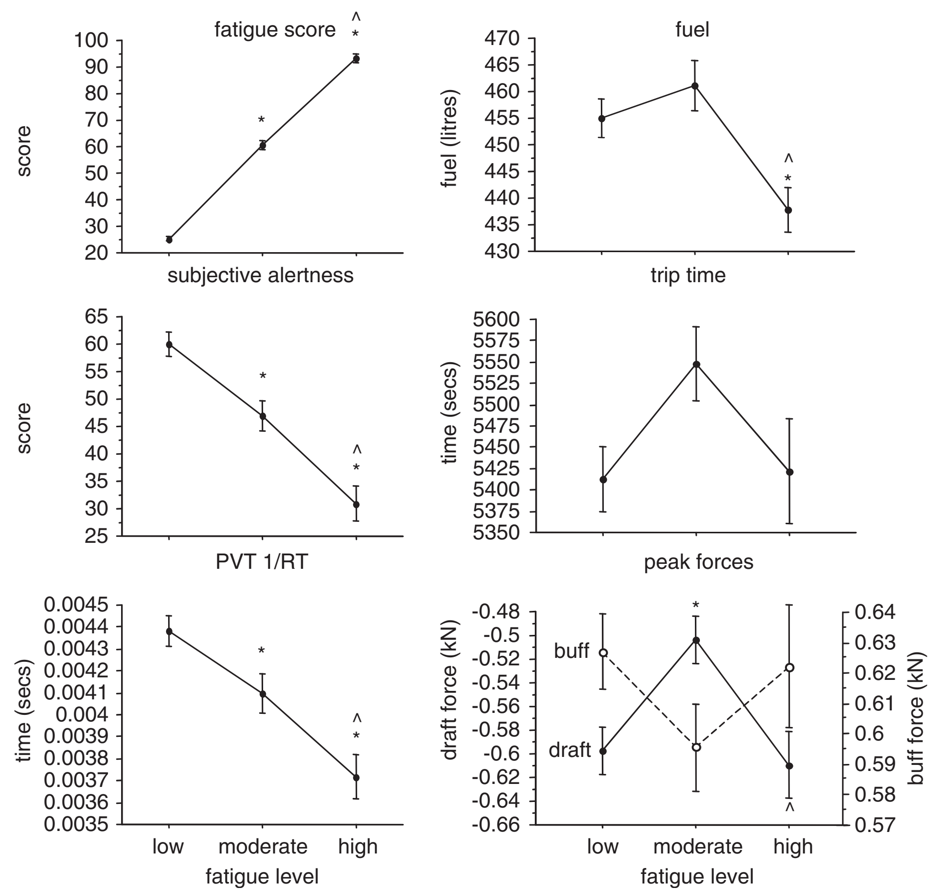

Fig. 2. Fatigue score, alertness, PVT, fuel use, trip time and forces at low, moderate and high fatigue levels. * indicates that scores are significantly different from those at a low fatigue level. ^ indicates that scores are significantly different from those at a moderate fatigue level.

subsequently decreased from moderate to high levels. In contrast, the number of extreme speed violations and penalty brake applications increased in a linear fashion with increasing fatigue level. Therefore, two main patterns were observed for driving errors; one pattern for braking errors, and another for extreme speed violations and penalty brake applications (see Fig. 5). It must be noted therefore, that while the pattern in braking errors (and subsequent fuel usage) would suggest a positive effect of fatigue on train driving performance, there is in fact a substantial decrease in driving safety, as indicated by the elevated speeding and penalty brake applications.

According to Reason's (1990) Generic Error Modelling System (GEMS), errors can be classified according to the cognitive level at which they are committed. Specifically, they fall into three categories: (i) skill-based slips and lapses, (ii) rule-based mistakes and (iii) knowledge-based mistakes. Slips and lapses typically occur during routine action sequences, resulting from a mistimed or omitted attentional check. Previous research indicates that train driver errors are mainly at the skill-based level (Edkins and Pollock, 1997). This is due to the task characteristics of train driving, which is a largely routine task that habitually involves travel over a previously practised, well-known stretch of track (Edkins and Pollock, 1997).

Based on this aspect of the GEMS (Reason, 1990), Fig. 6 illustrates a suggested map of driving error causation to explain the findings of this study. Specifically, fatigue-related limited attentional resources result in late, or neglected attentional checks. For this reason, the driver fails to plan adequately for an upcoming speed restriction such that they exceed, or are about to exceed the track limit. At this point, there are three potential scenarios. Firstly, the driver may apply the brakes late, but correctly (i). In this case the driver will make appropriate brake reductions, avoiding braking errors. A speed violation will still likely occur, although it will be somewhat minimised. Secondly, the driver may apply the brakes and in an effort to reduce train speed as quickly as possible, may make one or more braking errors (ii). While the speed violation is again somewhat minimised in this scenario, the error(s) will cause potentially damaging and dangerous drawbar forces along the train. Thirdly, the driver may fail to apply the brake (iii). Forseeably, this scenario could result from (a) a continued absence of attentional checks and thus a failure to realise the problem, or (b) decreased motivation due to increased fatigue and a lack of incentive to maintain driving standards. This scenario is particularly dangerous as it is likely to yield extreme speeding violations and significantly increase derailment risk. While scenarios (i)

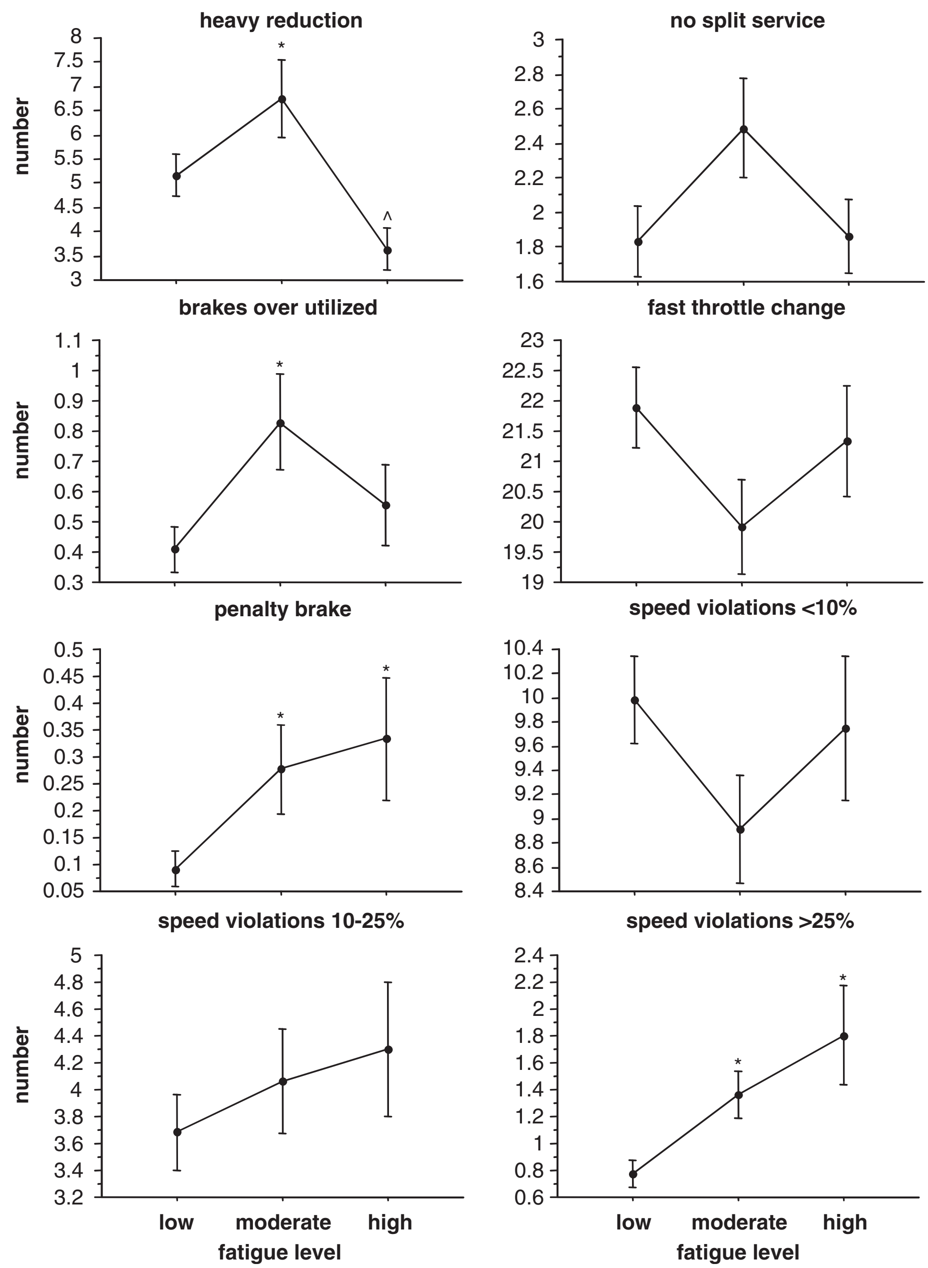

Fig. 3. Driving errors and violations at low, moderate and high fatigue levels. * indicates that scores are significantly different from those at a low fatigue level. ∧ indicates that scores are significantly different from those at a moderate fatigue level.

and (ii) involve acting to minimise the problem, scenario (iii) involves a failure to act.

The two patterns observed in error making in the current study (i.e. braking violation pattern, extreme speed violation pattern) fit this conceptual model. From low to moderate fatigue levels, there was an increase in the occurrence of scenario (ii), and thus an increase in braking errors (errors of commission). The resultant decrease in driving efficiency was reflected in the increase in fuel use and trip time. However, from moderate to high fatigue levels, there was an increase in the occurrence of scenario (iii), and thus a decrease in braking errors and a concurrent increase in extreme speed violations (errors of omission). Therefore, an increase in driving efficiency was suggested, reflected in the decrease in fuel use and trip time. Importantly however, as mentioned earlier, this increase in 'efficiency' was a false economy given the concomitant increase in safety risk.

Thus, from low to moderate fatigue levels, we observed a parallel increase in errors of commission and omission—consistent with 'state instability.' However the relationship changed from moderate to high fatigue levels. One possible explanation for this could be provided by a *cognitive-energetics* approach, which suggests that people engage in a type of energy/performance cost-benefit trade-off. That is, when confronted with a stressor, they can either recruit

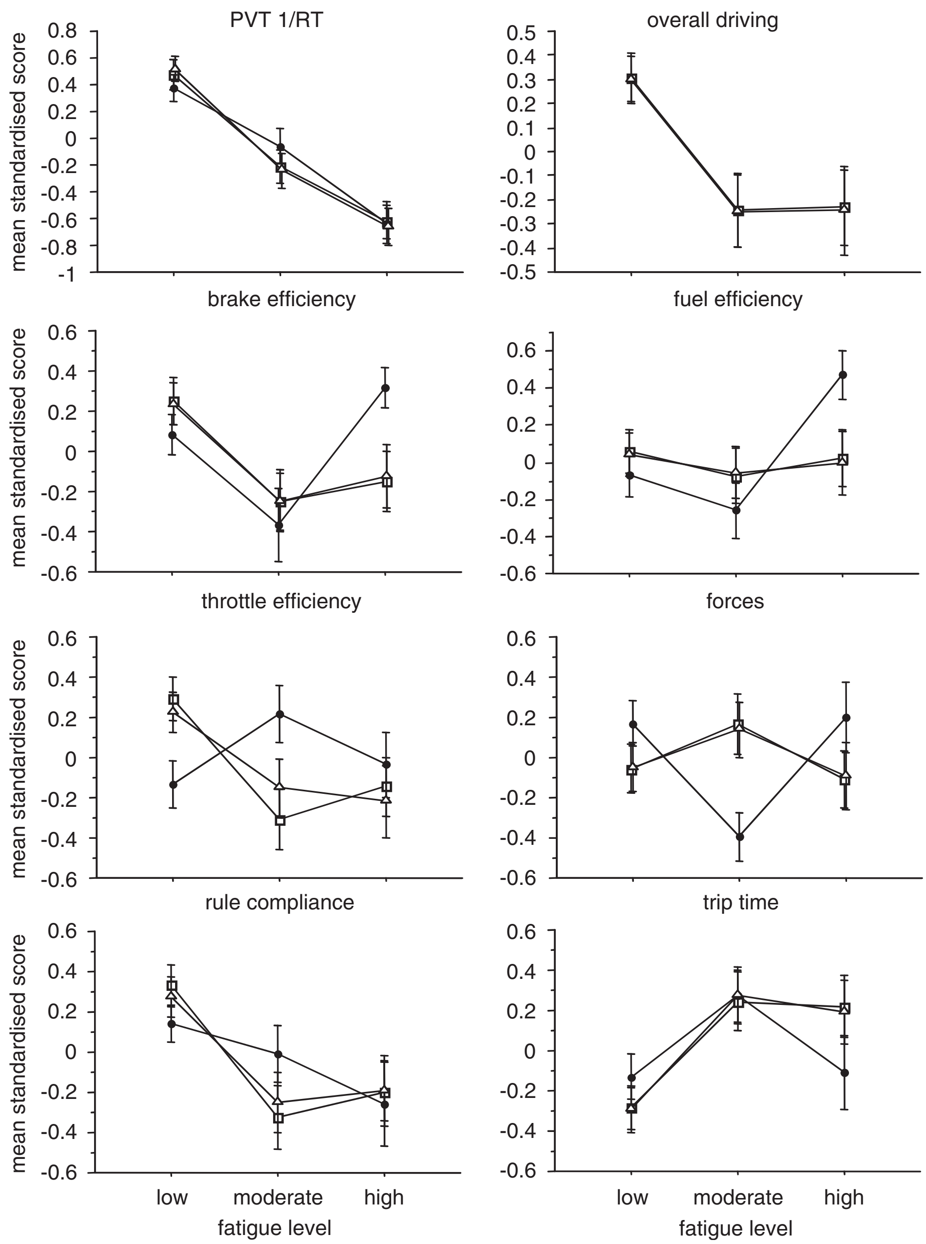

Fig. 4. Actual performance (●), pre-test (△) and post-test (□) ratings for PVT and driving measures.

extra resources and increase effort in order to attain performance goals, or they can reduce performance goals to avoid an increase in effort. Specifically, three patterns of response to stress, or coping modes have been proposed, where a person: (1) increases their effort in order achieve their performance goals, keeping effort expenditure within reserve limits—active coping; (2) increases their effort beyond reserve limits—strain coping; and (3) reduces performance goals in order to avoid over-spending their energy—passive coping. In its most acute form, passive coping results in a complete disengagement from the task at hand (Hockey, 1997).

In this light, results of this study suggest that increasing fatigue (placing increased stress on the driver) may result in a transition from active and strain coping modes, to passive coping. From low to moderate fatigue levels, drivers exhibited an increased number of braking errors. It could be suggested that they were entering a strain coping mode, where the cost was born out in reduced driving efficiency. However, between moderate and high levels of fatigue, drivers may have transitioned to a passive coping mode. That is, they may have become disengaged such that their interaction with the simulator declined dramatically and they 'let the train run itself.' In this way, errors of commission (e.g. braking errors) decreased, while errors of omission (e.g. speed violations) increased. In essence, a paradigm shift in behaviour was observed from impaired driving performance to *cognitive disengagement* between moderate and high fatigue levels. Further evidence for driver-train disengagement lies in the frequency of penalty brake applications (errors of omission), which also increased with fatigue.

Given previous work, which suggests that during conditions of elevated fatigue, an individual will maintain tasks that are most safety-critical at the expense of others (Fairclough and Graham, 1999; Hockey, 1997) we would have expected to observe minimisation of extreme speeding violations (i.e. violations most likely to cause derailment) at the expense of brake errors. Rather, the opposite was observed. However, the unreal environment of the simulator may have reduced the motivation of train drivers to drive safely, especially given that it was impossible to 'derail' our virtual train. In the real work environment, conservation of braking errors at the expense of safety, as observed in the current study, may be less likely to occur.

Indeed, the issue of motivation due to lack of consequences is one of the primary limitations of simulator studies. It could be suggested that in the real world, fatigue will still likely affect performance, but may manifest differently. For example, referring back to Fig. 5, fatigued drivers in a real train may be more likely to continue to apply the brakes late but incorrectly (option (ii)) in order to avoid dangerous speed violations. A further potential limitation of the current study is the fact that analyses investigated the track as a whole. While this was useful to gain an overall picture of driving performance changes with fatigue, it would be also beneficial to investigate where in the track the different driving errors occurred. This would allow analysis of the potential interaction effects between fatigue and track characteristics. In addition, it would provide context for the violations. For example, over utilisation of the brake (i.e. depleting the brake pipe pressure such that no further brake applications are possible) is more serious on a downhill section than on a flat section of track.

Interestingly, buff and draft forces were lowest at moderate fatigue levels. Given the pattern in braking errors, this result seems counter-intuitive. However, the magnitude of intertrain forces increases in proportion with train speed. Thus, since trip time was highest at a moderate fatigue level (indicating slower average train speed), the findings for intertrain forces are easily explained. Moreover, as our virtual train was short and heavy, and the track contained numerous speed restrictions and long uphill sections, the train speeds reached during this study were relatively low. Therefore, forces remained at quite low magnitudes, despite errors in brake and throttle manipulation.

Table 2
Mixed model results for pre- and post-test performance ratings [random effect: subjects, fixed effect: fatigue group, dependent variables: column 2].

		$F_{(2,\ 157)}$	P
Pre-test	PVT	31.80	>0.01
	Brake	5.99	<0.01
	Throttle	6.06	<0.01
	Fuel	0.90	NS
	Forces	0.76	NS
	Rules	5.43	<0.01
	Trip time	6.76	<0.01
	Overall	7.77	<0.01
Post-test	PVT	31.90	<0.01
	Brake	5.32	<0.01
	Throttle	8.21	<0.01
	Fuel	0.25	NS
	Forces	0.57	NS
	Rules	6.97	<0.01
	Trip time	6.26	<0.01
	Overall	7.14	<0.01

Table 3
Mixed model results comparing pre- and post-test performance ratings with actual performance and subjective alertness [random effect: subjects, fixed effect: pre- and post-test performance ratings, dependent variables: performance scores or alertness ratings]

		Pre-test			Post-test		
		$F_{(1,\ 158)}$	P	AIC[a]	$F_{(1,\ 158)}$	P	AIC[a]
VAS and Actual	PVT	139.87	<0.01	−1963.50	189.5	<0.01	−1985.0
	Brake	3.33	NS	828.76	16.48	<0.01	821.11
	Throttle	0.90	NS	611.09	0.98	NS	597.90
	Fuel	0.02	NS	1428.20	0.35	NS	1437.09
	Forces	0.72	NS	−8.87	0.46	NS	−9.38
	Rules	2.1	NS	1083.4	24.73	<0.01	1069.34
	Trip time	0.02	NS	3756.3	0.29	NS	3779.7
VAS and Alert	PVT	197.48	<0.01	1255.75	65.34	<0.01	1324.03
	Brake	64.87	<0.01	1351.15	35.71	<0.01	1366.27
	Throttle	61.79	<0.01	1354.63	35.33	<0.01	1366.31
	Fuel	15.97	<0.01	1392.46	8.91	<0.01	1390.63
	Forces	11.18	<0.01	1394.08	48.7	<0.05	1392.09
	Rules	63.07	<0.01	1352.67	17.13	<0.01	1383.27
	Trip time	21.67	<0.01	1384.63	9.34	<0.01	1386.88

[a]Akaike's Information Criterion (AIC), with smaller values indicating better model fit.

Table 4
Pearson *r* correlations between actual performance, alertness and pre- and post-test ratings

	Pre-test and actual	Post-test and actual	Pre-test and alertness	Post-test and alertness
PVT	0.54	0.50	0.72	0.54
Brake	0.39	0.36	0.54	0.46
Throttle	−0.08	−0.06	0.56	0.46
Fuel	0.19	0.21	0.35	0.28
Forces	−0.05	0.02	−0.25	0.18
Rules	0.19	0.32	0.54	0.36
Trip time	0.11	0.34	−0.34	−0.25

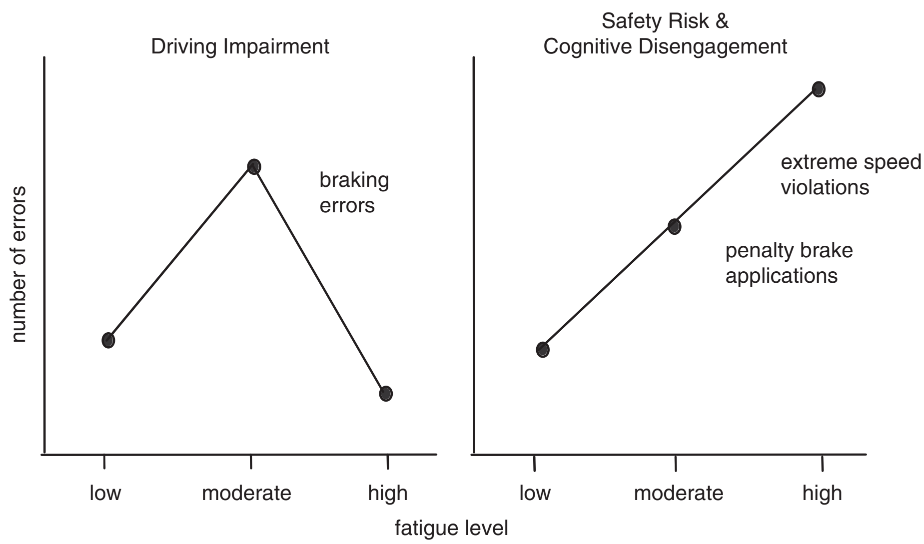

Fig. 5. Schematic representation of patterns in driving errors.

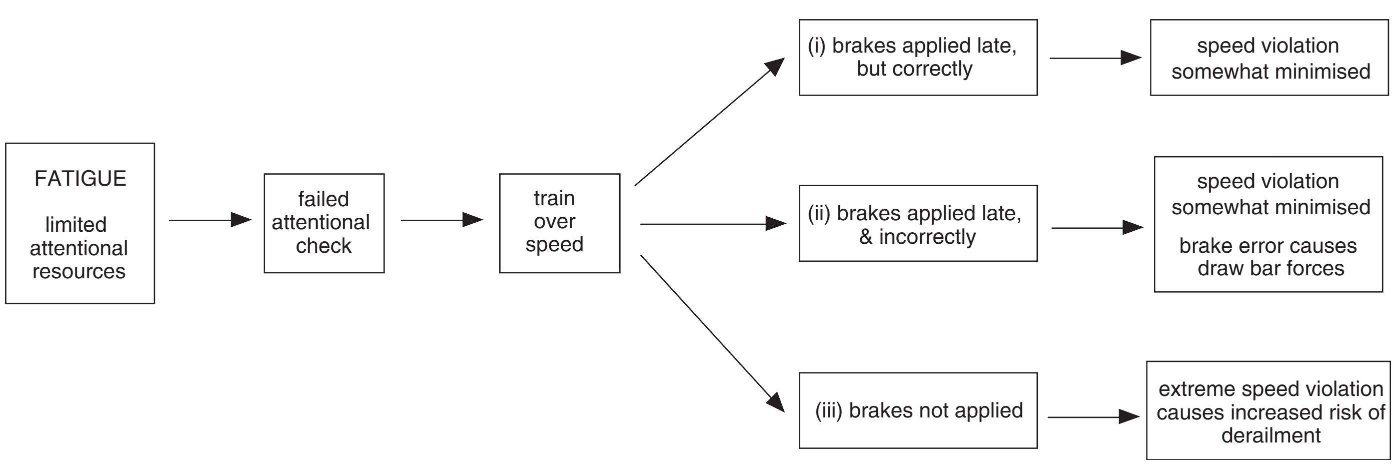

Fig. 6. Map of driving error causation.

4.2. *Self-monitoring performance*

Results were, to an extent, consistent with previous studies indicating that fatigued individuals are able to accurately rate performance impairment (Baranski et al., 1994; Baranski and Pigeau, 1997; Dorrian et al., 2000). Performance ratings were found to be significant predictors of performance for PVT (pre- and post-test), brake efficiency and rule violations (post-test only). Indeed, moderate correlation was found between ratings and actual performance for these parameters. Thus, it could be inferred that these ratings were reasonably accurate. However, taking into account the severity of fatigue-related rail accidents (e.g. described in Lauber and Kayten, 1988; Pearce, 1999), a relationship of this modest magnitude may still be considered operationally concerning.

The association found between self-ratings and performance for all other parameters was low at best. This was particularly the case for throttle efficiency and forces. In fact, ratings reflected a decrease in throttle efficiency with increasing fatigue level whereas, in general, an increase was observed. Similarly, ratings indicated highest forces at moderate fatigue, when the opposite was observed. However, variation in throttle efficiency was not significant and, as discussed earlier, forces incurred during this study were relatively low. Therefore, it may have been more difficult for drivers to appreciate changes in these parameters.

Ratings were more accurate for PVT performance than for any simulator measure, suggesting that performance on a simple laboratory test may be more easily appreciated than performance on a more complicated 'real world' task. Previous studies finding accurate performance ratings while fatigued (Baranski et al., 1994; Baranski and Pigeau, 1997; Dorrian et al., 2000), were conducted using laboratory tests of performance. In contrast, when performance in a driving simulator was measured, only a weak relationship between perceived and actual performance deterioration was found (Arnedt et al., 2000).

In general, post-test ratings resulted in better fitting models of actual performance than pre-test ratings. This is consistent with previous findings suggesting greater post-test/retrospective rating accuracy (Arnedt et al., 2000; Dorrian et al., 2000, 2003). Overall ratings of simulator performance indicated that subjects felt that their performance would become increasingly impaired as they became increasingly fatigued and less alert. This was reflected in their ratings for individual performance parameters. Consistent with previous research, a strong relationship was found between subjective alertness and pre- and post-test ratings for all aspects of performance (Dorrian et al., 2000). Interestingly, post-test ratings were poorer predictors of alertness than pre-test ratings. This suggests that post-test, when individuals had more information about their performance, they relied less on alertness as a cue for performance rating. This is consistent with previous work which suggested that alertness may mediate performance ratings to a larger extent in the absence of more direct performance feedback (Dorrian et al., 2003). Nevertheless, current findings suggest that subjects were aware that not all aspects of performance would be affected by fatigue in the same way. In a prior study by our research group investigating the ability accurately rate performance on six neurobehavioural tasks during 28-h awake, ratings for all parameters were equivalent despite the differential effect of fatigue on these parameters (Dorrian et al., 2000). In contrast, in the current study, while participants expected significant changes in brake and throttle efficiency, rules and trip time with increasing fatigue level, the same was not true for fuel efficiency and intertrain forces.

5. Conclusions

Taken together, findings of the current study have a number of implications for rail accident research. Results indicate that high levels of fatigue may result in a *cognitive disengagement* from the driving task, and thus a dramatic increase in accident risk. While individual ability to appreciate fatigue-related performance changes may be reasonable for certain aspects of rail operating, it may be negligible for others. Given the suggestion that safety-protecting responses to increasing fatigue are cued by self-awareness of rising performance impairment (Fairclough and Graham, 1999), these findings may be considered operationally concerning.

Acknowledgements

The authors would like to acknowledge Pat Wilson and Frank Hussey for their assistance with the technical aspects of train driving described in this manuscript. This research was supported by The Australian Shift Work and Work Load Study

References

Arnedt, J.T., Wilde, J.S., Munt, P.W., Maclean, A.W., 2000. Simulated driving performance following prolonged wakefulness and alcohol consumption: separate and combined contributions to impairment. J. Sleep Res. 9, 233–241.

Baranski, J., Pigeau, R., 1997. Self-monitoring cognitive performance during sleep derivation: effects of Modafinil, D-Amphetamine and placebo. J. Sleep Res. 6 (2), 84–91.

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

Caldwell, J.A., Smythe, N.K., Leduc, P.A., Caldwell, J.L., 2000. Efficacy of Dexedrine for maintaining aviator performance during 64 h of sustained wakefulness: a simulator study. Aviat. Space Environ. Med. 71 (1), 7–18.

Dawson, D., Fletcher, A., 2001. A quantitative model of work-related fatigue: background and definition. Ergonomics 44 (2), 144–163.

Dawson, D., Reid, K., 1997. Fatigue, alcohol and performance impairment. Nature 388 (6639), 235.

Dinges, D.F., Powell, J.W., 1985. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. Behav. Res. Methods Instrum. Comput. 17, 652–655.

Dinges, D.F., 1992. Probing the limits of functional capability: the effects of sleep loss on short-duration tasks. In: Broughton, R.J., Ogilivie, R.D. (Eds.), Sleep, Arousal and Performance. Birkhauser, Boston, MA.

Dinges, D.F., Powell, J.W., 1988. Sleepiness is more than lapsing. Sleep Res. 17, 84.

Dinges, D.F., Powell, J.W., 1989. Sleepiness impairs optimum response capability. Sleep Res. 18, 366.

Doran, S.M., Van Dongen, H.P.A., Dinges, D.F., 2001. Sustained attention performance during sleep deprivation: evidence of state instability. Arch. Ital. Biol. 139, 253–267.

Dorrian, J., Lamond, N., Dawson, D., 2000. The ability to self-monitor performance when fatigued. J. Sleep Res. 9 (2), 137–145.

Dorrian, J., Lamond, N., Holmes, A.L., Burgess, H.J., Roach, G.D., Fletcher, A., Dawson, D., 2003. The ability to self-monitor performance during a week of simulated night shifts. Sleep 26 (7), 871–877.

Dorrian, J., Rogers, N.L., Dinges, D.F., 2004. Behavioural alertness as assessed by psychomotor vigilance performance. In: Kushida, C. (Ed.), Sleep Deprivation: Clinical Issues, Pharmacology, and Sleep Loss Effects. Marcel Dekker, New York, pp. 39–70.

Edkins, G.D., Pollock, C.M., 1997. The influence of sustained attention on railway accidents. Accid. Anal. Prev. 29 (4), 533–539.

Fairclough, S.H., Graham, R., 1999. Impairment of driving performance caused by sleep deprivation or alcohol: a comparative study. Human Factors 41 (1), 118–128.

Fletcher, A., Dawson, D., 2001. A quantitative model of work-related fatigue: empirical evaluations. Ergonomics 44 (5), 475–488.

Foret, J., Latin, G., 1972. The sleep of train drivers: an example of the effects of irregular work hours on sleep. In: Colquhoun, W.P. (Ed.), Aspects of Human Efficiency. English Universities Press, London.

Gillberg, M., Kecklund, G., Åkerstedt, T., 1994. Relations between performance and subjective ratings of sleepiness during a night awake. Sleep 17 (3), 236–241.

Hockey, G.R.J., 1997. Compensatory control in the regulation of human performance under stress and high workload: a cognitive-energetical framework. Biol. Psychol. 45, 73–93.

Kjellberg, A., 1977. Sleep deprivation and some aspects of performance. I. Problems of arousal changes. Wak. Sleep. 1, 139–143.

Kleitman, N., 1963. Deprivation of sleep. In: Kleitman, N. (Ed.), Sleep and Wakefulness, pp. 215–229.

Kogi, K., Ohta, T., 1975. Incidence of near accidental drowsing in locomotive driving during a period of rotation. J. Human Ergol. 4 (1), 65–76.

Lauber, J.K., Kayten, P.J., 1988. Sleepiness, circadian dysrythmia, and fatigue in transportation accidents. Sleep 11 (6), 503–512.

Patrick, G.T., Gilbert, J.A., 1896. On the effects of loss of sleep. Psychol. Rev. 3, 469–483.

Pearce, K., 1999. Australian Railway Disasters. IPL Books, NSW, Australia.

Pilcher, J.J., Coplen, M.K., 2000. Work/rest cycles in railroad operations: effects of shorter than 24-h shift work schedules and on-call schedules on sleep. Ergonomics 43 (5), 573–588.

Pollard, J., 1991. Issues in Locomotive Crew Management and Scheduling. Federal Railroad Administration, US Department of Transportation, Washington, DC.

Pollard, J., 1996. Locomotive Engineer's Activity Diary. Federal Railroad Administration, US Department of Transportation, Washington, DC.

Reason, J.T., 1990. Human Error. Cambridge University Press, USA.

Roach, G.D., Reid, K.J., Dawson, D., 2003. The amount of sleep obtained by locomotive engineers: effects of break duration and time of break onset. Occup. Environ. Med. 60 (12), e17.

Roach, G.D., Fletcher, A., Dawson, D., 2004. A model to predict work-related fatigue based on hours of work. Aviat. Space Environ. Med. 75 (Suppl. 3), A61–A69 discussion A70–A704.

Roth, E., 2000. In: Reinach, S., Raslear, T. (Eds.), An Examination of Amtrak's Acela High Speed Rail Simulator for FRA Research Purposes. Federal Rail Administration, Washington, DC, 2001.

Tabachnick, B.G., Fidell, L.S., 1996. Using Multivariate Statistics. Harper Collins College Publishers, California State University, Northridge.

Thomas, G.R., Raslear, T.G., 1997. The Effects of Work Schedule on Train Handling Performance and Sleep of Locomotive Engineers: A Simulator Study. US Department of Transportation, Federal Railroad Administration, Washington, DC.

Williams, H.L., Lubin, A., Goodnow, J.J., 1959. Impaired performance with acute sleep loss. Psychol. Monogr. Gen. Appl. 73 (14), 1–25.

Zhou, D.S., 1991. Epidemiological features and causes of railway traffic accidents. Zhonghua Yu Fang Yi Xue Za Zhi 25 (1), 26–29.

Further reading

Bohlin, G., 1971. Monotonous stimulation, sleep onset and habituation of the orienting reaction. Electroencephalogr. Clin. Neurophysiol. 39, 145–155.