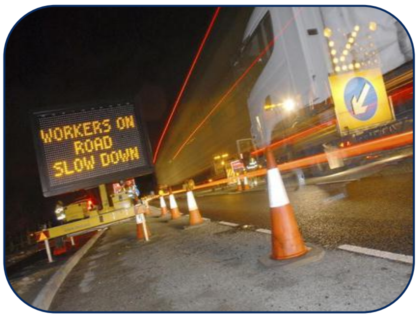
Transport Research Laboratory
Creating the future of transport




PUBLISHED PROJECT REPORT PPR748

A pilot study of low workload in train drivers

D. Robinson, S. Waters, D. Basacik, A. Whitmore and N. Reed

Prepared for: TRL-RSSB Joint Research Board

Project Ref: 11112576

Quality approved:

S Cotter (Project Manager) *S. Cotter*

A Parkes (Technical Referee) *A. Parkes*

Disclaimer

Contents amendment record

This report has been amended and issued as follows:

Version	**Date**	**Description**	**Editor**	**Technical Referee**
0.1	28/11/2014	First draft for internal review	SW	
0.2-0.5	02/12/2014	Redrafts following internal review	SW	
0.6	10/12/2014	Draft for technical review	SW	AP, A Mi
1.0	21/05/2015	Final draft	SW	
Final	28/05/2015	Final report		

Contents

1 Introduction 7

2 Literature review 9

2.1 Aims of the literature review 9

2.2 Method 9

2.3 Findings 9

2.3.1 Key terms and definitions 9

2.3.2 The effects of low workload on operator/driver performance 14

2.3.3 The methods of mitigating the negative effects of low workload 16

2.3.4 Methods of predicting and measuring low workload 19

2.4 Summary 21

3 Review of train driving tasks 23

3.1 Method 23

3.1.1 Analysis of the current GB train driving task 23

3.1.2 Task demand factors 23

3.1.3 Coding 25

3.1.4 Understanding the effects of ERTMS on workload 26

3.2 Outcome of the review of the current GB train driving task 26

3.2.1 Tasks that place lower demands on the driver 26

3.2.2 Key situational factors that can influence demand 27

3.2.3 Implications for the simulation trials 28

3.3 Implications of ERTMS on train driver workload 28

4 Identification and selection of mitigation 31

5 Simulation trials 33

5.1 Method 33

5.1.1 Participants 33

5.1.2 Experimental design 33

5.1.3 Materials 34

5.1.4 Procedure 38

5.1.5 Hypotheses 40

6 Results 41

6.1 Performance measures 41

6.1.1 Over and underspeed events 41

6.1.2 Speed variability 41

6.1.3 Mean speed 42
6.1.4 Track worker reaction times 43
6.1.5 Response to track workers 43
6.1.6 Response to AWS irregularities 44
6.1.7 Reaction time to level crossing failure 44
6.1.8 DSD pedal activations 45
6.2 Subjective measures 46
6.2.1 NASA TLX 46
6.2.2 Karolinska Sleepiness Scale (KSS) 47
6.2.3 Stress Arousal Checklist (SACL) 47
6.2.4 Interview response 50
6.3 Physiological measures 51
6.3.1 Heart rate 51
7 Discussion 52
8 Conclusions 56
9 References 57

Executive summary

There are many scenarios within the rail industry in which the train drivers monitor the status of the train and environment, but make few control inputs. Many services run on cautionary signals or at low speeds for extended periods of time, and some passenger services may involve frequent station stops. In addition, there is increasing automation in the rail industry, resulting in front line staff roles becoming more supervisory in nature. There is an emerging concern that the train driving task may not demand sufficient attention from the driver to keep them alert and engaged, raising the risk that safety-critical information may be missed or not acted upon in the most appropriate manner. Cognitive underload occurs when the demands of a task are so low that the performance of the person carrying out the task suffers.

This study set out to investigate the types of train driving tasks and situational factors that might lead to low workload and reduced driver performance. It also investigated the effect that a mitigation designed to increase workload had on train drivers experiencing a relatively undemanding train driving scenario. The mitigation chosen was a series of task-related questions that were administered verbally by the experimenter, at fixed time intervals. Ten participants were recruited to take part in the study. Participants were required to complete two drives on Southern Railway's train simulator: one baseline drive and one in which the mitigation was applied. Subjective and physiological measures were taken for both the baseline and mitigation drives, as well as performance measures such as speed and response to critical events.

The results of the experiment did show that the mitigation had an effect on the way participants felt. When compared with their baseline drive:

- Participants reported overall higher ratings for perceived workload on the NASA-TLX questionnaire after the mitigation drive. Significantly higher ratings were reported for the perceived performance and effort categories.
- Self-reported sleepiness showed a delayed onset in the mitigation drive, when compared with the baseline drive.
- Self-reported arousal levels (on the Stress Arousal Checklist) were higher after the mitigation drive.

The heart rate measures suffered from data quality issues, so analysis was not possible. One explanation for these issues may have been that the electrodes that were in contact with the participants' chests may have become dry over time. This would have resulted in poor electrical signal conductivity.

Given the effect the mitigation had on the way participants felt, it is interesting that no differences were detected in the objective performance measures. Possible reasons as to why this might have been the case are:

- Motivation and effort could have been acting as a buffer against the detrimental effects of underload.
- The route may not have been familiar enough; therefore, performance issues associated with good prior route knowledge would have been missed.
- The compromises made during the study design may have had a detrimental effect. A training simulator was used instead of a research simulator. This limited the extent to which the test scenario could be adapted to the needs of the study

and meant that performance measures such as reaction times were taken in far less granularity than had originally been hoped.

The results from the study seem to suggest that applying a mitigation technique increases workload during a monotonous train driving scenario and has a positive impact on subjective measures of low workload, low arousal and fatigue.

1 Introduction

Technological advances, a competitive marketplace and developments in working practices within the rail industry have led to an increase in the level of automation that train drivers experience. Some passenger services run for extended periods of time without stopping at stations or encountering cautionary aspects at signals. Freight services tend to have longer journeys and they may be held in possessions, in sidings/depots or at stop signals for long periods of time. In such situations, the driver monitors the status of the train and environment, which may be slow to change. Many services run on cautionary signals or at low speeds for extended periods of time, and some passenger services may involve frequent station stops. Such work is characterised by the frequent repetition of action sequences, which can become monotonous. Therefore, there is an emerging concern that, in some circumstances and despite the driver selection process including criteria that aims to recruit individuals who are able to maintain attention and vigilance, the train driving task may not demand sufficient attention from the driver to keep them alert and engaged, raising the risk that safety-critical information may be missed or not acted upon in the most appropriate manner.

Task demands are the objective level of work or difficulty associated with a particular task. Tasks with low demand can be carried out with less effort, cognition and attention from the operator. Low demand is associated with low workload. The workload construct considers the operator's capabilities in relation to the objective demands of the task (Gopher & Donchin, 1986, as cited in Dunn 2011). Low workload tends to occur when an operator is performing a task that requires little cognition, effort or attention and the operator has little motivation to invest cognition, effort and attention into performing the task. When the low workload is associated with impaired performance in related tasks, it is referred to as cognitive underload.

There has been strong evidence suggesting that low workload can lead to several undesirable outcomes in drivers. These include reductions in driver concentration and motivation (Straussberger & Schaeffer, 2007), decrements to driving performance (Thiffault & Bergeron, 2003;), difficulties coping with emergency situations (Young & Stanton, 2006), reduced situational awareness of the driver (Gershon, Ronen, Oron-Gilad & Shinar, 2009) and increased passive fatigue (Desmond & Hancock, 2001, as cited in Saxby, Matthews, Hitchcock & Warm, 2007).

Specific train driving tasks and situations that are associated with low workload are not yet fully understood, and there is little research available on practical countermeasures that could potentially mitigate its deleterious effects.

The aim of this study was to identify the types of task and situational factors that might lead to low workload and reduced driver performance during passenger and freight train driving. It also investigated potential mitigations for reduced performance resulting from low workload. This was achieved through four research tasks:

1. A review of literature on low workload to identify characteristics of tasks or situations that could result in reduced driver performance, and to explore potential mitigations. This is presented in Chapter 2.
2. A desk-based study to identify tasks and task characteristics relevant to train driving that could result in low workload. This is the subject of Chapter 3.

3. A high level evaluation of mitigation measures that might prevent low workload or its negative impact on driver performance is described in Chapter 4.
4. A simulator study to measure the effects of one example cognitive underload mitigation method in terms of driver performance and feedback using a simulation methodology. This is presented in Chapter 5.

The results of the study are brought together and discussed in Chapter 7 and our conclusions are presented in Chapter 8.

2 Literature review

2.1 Aims of the literature review

The objective of this literature review was to determine the extent of current knowledge on the topic of low workload and cognitive underload. The aims of this review were therefore to identify:

- The definitions of key terms.
- Task-related factors associated with low demand.
- The effects of low task demand on operator/driver performance.
- Potential methods of mitigating the negative effects of low workload.
- Methods that could be used to predict and measure low workload.

The findings of the review are presented below and are structured to reflect the individual aims of the review.

2.2 Method

RSSB had access to Dunn's (2011) PhD thesis, *The Effect of Task Demand on Subjective Experience and Performance*. This thesis aimed to investigate the independent and interactive effects of task demand, monotony and fatigue on the driving performance of a train driver under a workload framework. It also contained a comprehensive review of the history and development of the scientific literature on workload. Dunn's thesis provides a theoretical backbone for this review. Additional pieces of literature were also reviewed to reflect developments since Dunn's thesis was published, to supplement the findings and provide more information to satisfy the objectives of this review.

For the additional literature searches two academic databases were used: Google Scholar and APApsychNET. Key terms associated with low workload (e.g. underload, workload) were used to search the databases for relevant entries. The abstracts of all the positive results were read. If the abstract contributed to the aims of the review, the paper was included. Literature that had already been reviewed by Dunn (2011) or that were not in accordance with aims of the review were excluded.

2.3 Findings

2.3.1 Key terms and definitions

Dunn (2011) suggests that to understand the concept of low workload and performance, several topic areas must be considered and understood:

- Arousal and alertness
- Attention and effort
- Task-related factors
    - Monotony
    - Task demands/workload
    - Time-on-task

- Individual factors
    - Sleep-related fatigue
    - Circadian rhythms
- Subjectively experienced states
    - Boredom
    - Fatigue (tiredness and sleepiness)
- Automation

These key terms will be briefly explained below.

From a physiological perspective, **arousal** is the level of central nervous system activation or excitation along a continuum of behavioural states (Razmjou, 1996, as cited in Dunn, 2011). These states can range from a *deep sleep* to a *state of alarm* (Grandjean, 1979, as cited in Dunn, 2011). Low workload situations are likely to be related to physiological states associated with drowsiness, weariness, or when the individual is hardly awake, relaxed or resting (Dunn, 2011). Theories of arousal also take into account the self-reported experiential feelings of the individual. Thayer (1978, as cited in Dunn, 2011) separated arousal into two phenomenological dimensions; '*energetic arousal*' and '*tense arousal*'. Energetic arousal is associated with feelings such as vigour, sleepiness and drowsiness. Tense arousal is associated with feelings such as tension, anxiety and placidity. Similar to the concept of arousal is **alertness** which is the conscious awareness of environmental and internal stimuli that allows an individual to concentrate on the task at hand (Shapiro et al. 2006). If arousal is defined as the *general* excitability of an individual then alertness can be described as the *task-directed* arousal (Dunn, 2011).

Dunn (2011) suggested arousal underpins the concept of workload in several ways: arousal is related to performance, arousal is related with attention, and arousal motivates and influences behaviour. The well-established Yerkes-Dodson Law states that if arousal increases, then task performance will improve up to an optimal point. After this point, further increases in arousal will result in decrements to performance. A fundamental assumption of workload theory posits that workload is strongly related to arousal: increases in workload will improve performance, but only up to an optimal point, after which further increases in workload will result in performance decrements. In regards to the Yerkes Dodson Law, workload and arousal are analogous and share a similar relationship with performance.

Research has consistently found arousal to be associated with attention (Dunn, 2011). Easterbrook (1959, as cited in Dunn, 2011) first noted that changes in arousal relate to changes in the range of environmental stimuli that is attended to. This is generally referred to as the '*attentional spotlight*'. In general, if arousal decreases, then the attentional spotlight starts to fall on a much wider area, resulting in a loss of focus.

Dunn (2011) also pointed out how arousal is linked with motivating behaviour. In general, higher arousal is linked with higher motivation to perform a given task. This is important for low workload/arousal situations as the individual will have lowered motivation to perform the task. This could lead to decrements in task performance or withdrawal from the task.

Attention is generally defined as the 'selective processing of incoming sensory information' (Driver, 2001). Contemporary models of attention are generally based upon Kahneman's (1973) limited capacity resource model. This model posits that an individual has limited attentional resources to cope with the demands of a given task. If the two levels match then the task will be carried out to an optimum level. If the demands of the task exceed the available resources then overload will occur and the individual's ability to attend to the environment will be impeded and decrements in performance are likely to result. From Kahneman's attentional resource theory it is not clear how low workload could lead to similar decrements in performance. Young and Stanton (2002a) proposed their malleable attentional resource theory (MART) to explain degradations in performance under low workload situations. Previous attentional resource theories assumed that the size of the pool of available attentional resources remained constant. MART posits that the size of an individual's pool of available attentional resources will adapt to suit the demands of the task. Under low workload conditions an individual's pool of available attentional resources will shrink to accommodate the lesser demands of the task. As there are fewer available attentional resources to carry out the task, the individual's ability to cope with the demands of the task will lessen and decrements in performance will begin to occur.

Generally, individuals will make subjective appraisals of the demands of the task and allocate attentional resources proportionally (Fairclough, 2001, as cited in Dunn, 2011). This view emphasises the *active* role of the individual in regulating attention. However, Lavie's perceptual load theory emphasises a more *passive* role for the individual. This theory posits that an individual's level of attention can also be influenced by the quantity of environmental information (Lavie, 1995, 2005). In low workload situations, it is likely that the individual will have a lowered state of attention. This is because there is likely to be less environmental information for the individual to process and when they make subjective appraisals of the demands of the task they are likely to underestimate the level of attentional resources that are required in order to cope with the actual demands of the task (Matthews and Desmond, 2002). Consequently, this is likely to lead to decrements in performance.

Individuals have a degree of control over their ability to attend to environmental stimuli but the perceptual demands of the environment will also influence attention. Resource theories of attention tend to agree that attentional resources are allocated to a given task. Individuals determine the level of attentional resources that are allocated to a given task based on subjective evaluations of the given task. Tasks evaluated to have low demands are allocated fewer attentional resources.

The allocation of attentional resources is analogous with the more intuitively experienced concept of mental effort (Dunn, 2011). Mental effort is the effortful attention that is applied to a given task. In general, Hockey's (1997) compensatory control model suggests mental effort can be applied to most tasks without conscious effort and in low workload tasks the mental effort is reduced with the intention of conservation. This lowered mental effort is likely to lead to performance decrements.

Vigilance or **sustained attention** can generally be regarded as two similar constructs. Oken, Salinsky and Elsas (2006) reviewed the research on sustained attention and vigilance (also referred to as tonic alertness). They concluded that idiosyncratic methodological differences between earlier 'vigilance' research and more recent 'sustained attention' research may have given rise to the discrepancies among existing definitions but all definitions seem to share an underlying concept: that they are associated with states of sufficient alertness to monitor one's environment effectively, with a particular emphasis on scanning for potentially dangerous stimuli. One's level of alertness is associated with one's more general attentional capacity (Sarter, Givens & Bruno, 2001) and reductions in sustained attention have been shown to negatively influence performance.

The **task-related factors** associated with low workload situations are the objective characteristics of a given task that lead to underload in the individual. Dunn (2011) suggests there are three task-related factors that have been shown to be strongly associated with low workload situations; monotony, time-on-task, and the objective demands of the task.

There is no agreed definition for **monotony** and previous research has used terms such as "monotonous state", "monotonous working conditions", "underload" and "boredom" interchangeably leading to mixed results and conclusions about the effects of monotony on workload and performance (Dunn, 2011). McBain's (1970) distinction between objective monotony and subjective monotony is adopted by Dunn. Objective monotony encompasses the objective conditions of a particular task, such as repetition and the predictability of the tasks and is defined as an 'objective task-related characteristic of an environment that is unchanging or that changes in a repetitive and predictable way' (Dunn, 2011). Subjective monotony is the subjective appraisal of a particular task or tasks and is usually consciously manifested and experienced as boredom. The distinction is important as monotonous tasks do not necessarily lead to boredom and boredom states can arise from non-monotonous tasks.

The **time-on-task** is the 'length of time an individual has been performing a particular task and may also be referred to as time on duty, time onto work shift, or driving time' (Dunn, 2011). It has been well-documented that the time-on-task is positively associated with degradation of performance although Dunn (2011) highlights limitations of much research in this field as it consistently fails to adequately control for potential confounding variables, such as the time of day, frequency/length of rest breaks, or roster patterns.

The **objective demands of the task** are the objective level of work or difficulty associated with the task. This is also known as the objective workload of the task. Traditionally, task demands have been categorised according to Hart and Staveland's (1988) NASA-TLX: perceptual demands, mental demands and physical demands. Thus low perceptual demands would be typified as stimuli that are difficult to see or hear. Low mental demands would include a lack of a requirement to think, plan, make decisions, switch attention or remember information. Low physical demands are related to lack of movements, high levels of comfort or inability to walk around.

More recently, MacDonald (2003) conducted a review of the literature and identified five groups that task demands could be classified as. These suggested demand groups are:

- Sensory/perceptual demands – synonymous with Hart and Staveland's (1988) perceptual demands.
- Central processing demands – synonymous with Hart and Staveland's (1988) mental demands.
- Psychomotor demands – demands that require response accuracy, or specifically timed responses.

- Physical demands – synonymous with Hart and Staveland's (1988) physical demands.
- Affective demands - demands that require social engagement or emotion from the operator.

Monotony, time-on-task and objective task demands are three variables of significant interest to the current study and are considered during the design of the experimental scenario. Dunn (2011) suggested several other factors that are associated with low workload situations. These factors fall beyond the scope of this project was to focus on the task-related factors associated with low workload. However, they will be important to measure or control, and are therefore briefly discussed below.

The **individual factors** associated with low workload include sleep-related fatigue and circadian rhythms. Broadly speaking, lack of sleep enhances the effects of low workload situations as it can reduce arousal levels. Williamson, Lombardi, Folkard, Stutts, Courtney and Connor (2011) conducted a review on fatigue and performance and defined **sleep-related fatigue** as fatigue caused by a 'reduction in the quantity or quality of sleep, or extension of the time awake since sleep' and this produces 'a homeostatic drive to sleep', which is often experienced as sleepiness.

Circadian rhythms are the 24 hour sleep-awake cycle that humans experience. Circadian rhythms have been well-documented to be linked with states of arousal and alertness (Dunn, 2011). Williamson et al (2011) concluded circadian rhythms are linked to safety outcomes. As with fatigue, the direct effects of circadian rhythms are difficult to empirically observe as they interact with other factors such as time since waking, rest breaks or lighting conditions.

Importantly, for the current simulator study, these findings strongly suggest these individual factors should be controlled. Instructions should be given to participants that suggest they have a normal night's sleep prior to the study. All trials should start at the same time of day. These control measures should reduce the possible confounding effects of sleep-related fatigue and circadian rhythms.

Dunn (2011) suggests the link between degradation of performance and the individual and situational factors outlined above occurs through two subjectively experienced states: boredom and fatigue. These two subjective states act as causal mechanisms between the antecedent factors of low workload and lowered arousal/performance degradation.

There is no universal definition for **boredom**, but all definitions appear to share a common feature. It is the negative phenomenological state experienced by an individual when they interact with an environment that they perceive to be non-stimulating (Fahlman, Mercer-Lyon, Flora & Eastwood, 2013). These environments tend to be unchanging, under-stimulating and constraining (Dunn, 2011), which are typical characteristics linked with low demand; therefore, boredom is likely to be associated with low workload situations. Boredom seems to negatively affect performance in several ways. It is directly and negatively associated with arousal, it can reduce task-directed attention and it can lead to distraction due to the individual's boredom-relieving strategies (Dunn, 2011).

Fatigue lacks an agreed definition throughout the literature, possibly because it is an 'umbrella' term that subsumes multiple variations or dimensions of fatigue (Hitchcock and Matthews, 2005, as cited in Saxby, Matthews, Hitchcock & Warm, 2005). Fatigue can be broadly conceptualised as the intermediate stages between being awake and sleep (Lai & Craig, 2001). Increased fatigue is associated with reductions in the efficiency to work (Grandjean, 1979, as cited in Lai & Craig, 2001) and an aversion to carry on work (Brown, 1994, as cited in Lai & Craig, 2001).

Two types of fatigue may be particularly associated with low workload: mental fatigue and passive fatigue (Desmond & Hancock, 2001, as cited in Saxby, Matthews, Hitchcock & Warm, 2007). Mental fatigue is associated with degradation of arousal, effort and performance. Passive fatigue results from prolonged task duration in which the operator is required to perform very few actions. Mental fatigue and passive fatigue are both likely to increase in low workload situations and this increase in fatigue is likely to result in performance degradation (Dunn, 2011).

Dunn (2011) highlights the importance of subjectively experienced states by suggesting that fatigue-associated subjective states provide a causal pathway between individual/situational factors that induce low workload and performance degradation. These two fatigue-associated subjective states are sleepiness and tiredness. Sleepiness seems to be a phenomenological manifestation of a mentally-fatigued state. A precise definition of sleepiness has remained elusive but Curcio, Casagrande & Bertini (2001) defined it as the state in which there is an increased propensity for the individual to fall asleep. Among the research and definitions there is a tendency to define sleepiness in terms of a state, but some research has investigated differences in trait sleepiness with some success. Trait sleepiness can be described as the characteristics of an individual that influence their experience of sleepiness in a given situation, or "an individual's starting point from where state sleepiness exerts its influence" (De Valck and Cluydts, 2003). However, it would seem reasonable that low workload situations are mostly associated with state sleepiness. Tiredness seems to be another subjectively experienced state induced by fatigue. A precise definition for tiredness seems equally elusive but Porr, Olsen and Hegadoren (2010) described tiredness as a normative response to stressors in everyday life and is subjectively experienced as an awareness of the gradual and predictable degradation of bodily functions. Overall, increases in sleepiness and tiredness are linked to decrements in performance (Dunn, 2011).

Automation has also been shown to be linked with low workload situations (Dunn, 2011). Automation is associated with advances in technology that have led to operating systems that require less action-based input from the operator. This has removed the operator from an active operating role and repositioned the operator in a more passive monitoring role (e.g. Parasuraman et al, 1993). Overall, empirical evidence on the effect of automation on driving performance and workload has been mixed (Merat, Jamson, Lai & Carsten, 2012). It seems likely that automation does lead to performance degradation but its relationships with low workload and performance are still not understood (Dunn, 2011).

2.3.2 The effects of low workload on operator/driver performance

Low workload seems to have several effects on the individual that act as a pathway to effects on performance. Low workload reduces the individual's arousal levels (Larue, Rakotonirainy & Pettitt, 2011) and their attention levels (Damrad-Frye and Laird, 1989; Sarter, Givens & Bruno, 2001). It is this lowered arousal and attention that plays a key part in a range of other negative effects on performance.

Dunn's (2011) model of low workload posits that there are four main effects of low workload on human performance. These effects have been well-documented throughout the literature. These effects are:

- Decreased performance on vigilance tasks.
- Increased reaction times.
- An increase in the number of performance errors.
- An increased chance of failing to respond to environmental stimuli.

In vigilance tasks, low workload has been shown to reduce the overall detection rate of critical stimuli over the duration of a task (Deaton & Parasuraman, 1997). This effect is only observed under specific conditions related to a vigilance task. However, more general effects have also been observed and these can apply to a wide range of tasks. Low workload increases the reaction times of the operator when they are responding to their environmental stimuli and cues. Low workload also increases the likelihood that the operator will fail to respond to the environmental stimuli. Furthermore, low workload will also increase the likelihood that the operator will make an error whilst carrying out a particular task.

Research that has focused on performance effects on car drivers in low workload situations have shown decrements in an individual's general driving performance (Gershon, Ronen, Oron-Gilad & Shinar, 2009). Thiffault and Bergeron (2003) found that participants made fewer steering adjustments in low workload scenarios, which suggests the drivers were paying less attention to the driving task. Young and Stanton (2006) found drivers had slower reaction times when responding to an emergency situation under low workload conditions and this suggests that drivers under low workload conditions may be less able to cope with emergency situations. Straussberger and Schaeffer (2007) found that drivers under low workload conditions can experience lowered levels of concentration, attention and motivation.

Train drivers may be particularly prone to experiencing low task demands. This is because train drivers do not steer the train and in some situations the potentially repetitive nature of accelerating and braking a train (Dunn, 2011). They are also likely to experience prolonged working hours combined with monotonous environmental driving routes (Edkins & Pollock, 1997). Train drivers themselves have recognised the problems of low task demand as monotony and fatigue have been reported as common issues within their working lives (Dunn & Williamson 2011). It is not yet fully understood what specific tasks of a train driver's role are associated with low workload. This review of literature has not found a robust analysis of the GB train driver's role to identify which tasks are associated with low workload.

Problems associated with low task demands such as maintaining attention and remaining vigilant may be the underlying causes for many rail incidents (Wilde & Stinson, 1983; Edkins & Pollock, 1997). Nachreiner (1995) argued that the effects of low task demands may be as hazardous as the effects of high task demands and this is likely to be exacerbated by the fact that low task demands are more difficult to detect (Young & Stanton, 2002b). There are few studies that focus on how low workload may relate to real-life safety outcomes. However, Williamson et al (2011) conducted a review and found that task-related fatigue and mental fatigue, which are related to low workload, are likely to be significantly involved with a minority of car crashes.

2.3.3 The methods of mitigating the negative effects of low workload

A brief review was conducted of several recent studies that investigated possible techniques for mitigating or preventing the negative effects of low workload on performance. The review included some papers on fatigue, particularly those which were looking at the effects of varying task demands on fatigue, as the workload elements of these studies were considered to be relevant to the current study.

Whilst a substantial volume of research has focused on situational factors that lead to underload and the subsequent effects on performance, much less research has focused on practical countermeasures that could potentially mitigate the deleterious effects of low workload.

Dunn and Williamson (2012) hypothesised that increasing the objective task demands of a driver under low workload conditions can prevent the deleterious performance effects associated with low workload. Theories suggest that low workload can reduce the individual's ability to attend to their environment (Young and Stanton, 2002a), reduce the individual's ability to maintain an optimum level of arousal (Hancock and Warm, 1989) and make the individual more likely to underestimate the level of effort needed to perform the task to a satisfactory level (Matthews and Desmond, 2002). According to these theories, if the driver is given extra cognitive load then the driver's ability to attend to the environment will improve, they will be more able to maintain an optimum level of arousal and they will be more likely to invest a more appropriate level of effort into the task to perform the task at a satisfactory level. Several studies have suggested that additional cognitive load whilst driving under low workload conditions can improve driving performance.

Three studies investigated the effects of an interactive secondary task (IST) on car driver performance. Verbal tasks could be used to mitigate the deleterious effects of low workload. Chan and Atchley (2011) conducted a simulator study that audibly presented drivers with a story whilst they were driving. The drivers had to listen and then retell the story. When drivers performed this verbal task their driving remained more steady (less deviation from the middle of the road) but they kept a shorter distance from the car in front. This indicates the verbal task can increase task-directed effort and performance but it could reduce risk/safety awareness.

Gershon, Ronen, Oron-Gilad and Shinar (2009) used an interactive cognitive task (ICT) in several car driving simulator studies to reduce driver fatigue and increase driver performance. The ICT was a series of multiple choice trivia questions presented to the driver though rear-positioned speakers. There were two important components of the ICT. The driver could select what theme the trivia questions were about and there were three levels of difficulty in the questions. The driver would start a new theme with the easiest question and move up a level if the question was answered correctly. This kept the driver interested and engaged with the task. The intention behind the variability in question difficulty was to increase task demands in underload situations but not to increase task demands to a level that would induce an overload situation. Gershon et al. (2009) found the ICT did reduce both subjective levels of fatigue and physiological measures of fatigue. The ICT also improved driving performance. Importantly, the effects of the ICT were only noticeable during the intervention itself. No long-terms effects were observed.

Verwey and Zaidel (1999) observed the car driving behaviour of two groups in a simulator. Both groups had to drive along the same monotonous route but one group had a 'gamebox' with which they were allowed to interact at will. The gamebox was a verbally interactive and mentally stimulating game that could be played by the drivers. The gamebox group reported feeling less drowsy than the control group and the gamebox group showed fewer physiological signs of drowsiness. Furthermore, drivers in the gamebox group made significantly fewer mistakes whilst driving. In both groups, there was still a consistent time-on-task effect that reduced overall performance and increased fatigue levels over time. Importantly, the drivers used the gamebox as they wished indicating that driver-controlled interventions against the negative effects of fatigue can also be an effective countermeasure.

It seems that ISTs can increase the demands of a task and improve driving performance by increasing attention, task-directed mental effort, arousal and alertness. Problems may arise when these ISTs interfere with other cognitive processes or capabilities such as risk awareness. The level of interactivity could be a particular strength of this type of mitigation as it allows the driver to choose the level of attention and effort that is applied to the secondary task. This provides some flexibility in how the driver responds to the changing demands of the driving task. If more effort is needed for the driving task, then less effort can be applied to the IST. These ISTs seem to be well-suited for improving the negative feelings of boredom and may also reduce fatigue. There is no evidence that they neither mitigate the time-on-task effects nor are their effects shown to be long term. Any positive effects of the mitigations seem to disappear as the ISTs are removed. ISTs may also pose a potential for distraction and may reduce risk awareness. From this, it seems these mitigations should be investigated in emergency situations or safety critical situations to investigate whether they impair performance.

Listening to the radio and talking on a mobile phone have received some attention within the literature as possible techniques for mitigating the negative performance effects associated with low workload. Anecdotally, they are frequently cited as methods for improving attention, effort and performance in car driving situations.

Listening to the radio may increase the demand for cognitive resources in low workload situations and improve driving performance. Ūnal, Platteel, Steg and Epstude (2012) conducted a simulator study that investigated the effects of listening to an in-car radio on driving performances. The study focused on the effects of listening to the radio in low and high traffic situations. Driving performance in the radio condition was significantly better than in the control condition. Post hoc interviews revealed that less information from the radio was recalled in the high traffic condition. This suggests car drivers are able to listen to music to improve performance in low workload situations and they can regulate their attentional resources to prioritise driving over listening to the radio.

Listening to the radio seems to increase the attentional resources available and these resources are allocated to the car driving task resulting in improved driving performance. Car drivers' ability to remember different levels of information depending upon the difficulty of the driving tasks supports the idea that the optimum level of resources are easily allocated to the driving task and that drivers are able to divert attention away from the radio if the driving task demands more effort. This is important as it suggests this mitigation does not compete with attentional resources that are required by the car driving task. Like the IST, however, further evidence should be sought in order to determine how listening to the radio may affect driving performance if the driver suddenly encounters a high workload or an emergency situation. In particular, listening to the radio may interfere with a driver's ability to attend to other audible stimuli relevant to driving. For example, the radio could compete with the audible warning alarms for the available attentional resources. With several audible in-cab warning alarms on trains, this may not be a practical solution.

Talking on a mobile phone whilst driving a car along a monotonous route can increase arousal and effort, which can lead to an improvement in driver performance. Jellentrup, Metz and Rothe (2011) investigated the changes in some of the physiological signs of fatigue (e.g. EEG recordings, eyelid closure level) and subjective ratings of fatigue whilst driving and conversing on a mobile phone. The participants drove a car around a long monotonous track for an extended period of time. In one condition they were provided with a mobile phone and they received a phone call during the drive. The physiological recordings indicated that drivers' fatigue was reduced during and after the mobile phone call. This was supported by the drivers' subjective ratings of their own fatigue, as the phone calls were rated as having a 'vitalizing effect' on their subjective states. However, Jellentrup, Metz and Rothe (2011) highlighted some important limitations of this study. The mobile phone calls were only effective in monotonous conditions as they were not effective when the driver was at a particularly low level of fatigue. The phone call intervention did not show long-term effects of reducing fatigue and the size of the effect of the mobile phone intervention lessened as the drivers became more accustomed to the mobile phone call conversation. This suggests it could be the novelty of the conversation that raises arousal rather than an increase in task demands.

Talking on the mobile phone whilst driving a car has produced some mixed results on driving performance (Saxby, 2012). There seems to be some evidence that it can improve driving performance on monotonous roads and this is supported by Jellentrup, Metz and Rothe (2011) which suggests mobile phone conversations may improve performance but only in conditions in which monotony is the significant factor. If fatigue is so low that it becomes the significant factor for low workload, then the positive effects of the mobile phone conversation are negated. Saxby (2012) suggests there are other factors associated with mobile phone use that may prevent it from being a viable mitigation method. These include reductions in situational awareness, sources of distractions and reducing the drivers' understanding of their own driving capabilities. Jellentrup, Metz and Rothe (2011) also highlight a potential problem of driver habituation to the positive effects of the mobile phone. This undermines the potential for mobile phones to be a potential long-term mitigation method on the railway as drivers would become accustomed to using them in their cab, thus negating their positive effects.

Two studies investigated environmental factors that could mitigate the negative effects of low workload. Infrastructure design may mitigate the effects of underload. Three types of engineering design solution were used in a simulator study and their effects on car driving performance and task-related fatigue was assessed (Merat & Jamson, 2013). The three engineering interventions were chevrons, rumble strips and variable message signs. All three were found to reduce the onset of task-related fatigue and maintain driving performance across the driving task. Chevrons denote safe distances between cars and rumble strips provide an auditory and vibratory cue to inadvertent lane drifting. Consequently, these are not generally applicable to the rail industry; however, the third option, variable message signs, could be considered. In the study, variable message signs, placed intermittently throughout the driving route, helped maintain the alertness of the car driver, increasing arousal and task effort. Similar signs could be used in the rail industry, although consideration would need to be given to potential infrastructure and operational issues as well as sign clutter.

The operator's immediate environment has also been shown to affect the performance of the operator under low workload conditions. Thornburg, Peterse and Liu (2012) assessed several operators' abilities to operate a mock computer-based nuclear safety control simulator. This simulator was designed to be monitored for extended periods (low workload) and intermittently interrupted by alarms that needed the operators' actions to reset the alarm or make the situation 'safe' again (high workload). There were two environmental conditions. The first 'sterile' condition had no external distractions other than the simulator itself. The second 'non-sterile' condition had other distractions the operator could use, e.g. a mobile phone, the internet or a book. The authors found two main effects of the 'non-sterile' environment. Operator performance was impaired by the presence of environmental distractors. Secondly, the length of time the operator was passively monitoring the simulator also had an effect. Longer monitoring times prior to an alarm being activated led to further impairments in performance when the operator had to deal with the alarm situation. Thornburg, Peterse and Liu (2012) noted that this effect was most prominent for monitoring times that lasted longer than an hour.

These two studies suggest that the environment can be used to increase arousal and task engagement but the environment can also be a source of distraction. There may be a potential optimum level of environmental stimuli that can be used to engage with the driver to increase driver performance, after which any increases in environmental stimuli will result in distraction or reduce task-related effort. There are several advantages of using the variable message signs. The variable messages on the sign are less likely to induce habituation in the driver, resulting in similar strength of responses towards the many signs. Secondly, the signs could be placed along long, monotonous routes that are likely to induce periods of low workload on the driver.

In-cab design should also be considered when mitigating low workload. Potential non-task-relevant distractions in the cab may be more of a problem in situations of low workload as they tend to be ignored in higher workload situations. Thornburg, Peterse and Liu's (2012) findings may also present a problem for other types of mitigation devices such as the previously described gamebox (Verwey and Zaidel, 1999) as these devices have the inherent possibility of becoming a distraction.

Overall, ISTs seem to have a positive effect on the car drivers' attention, task-related effort, performance, fatigue and boredom. In particular, the ISTs flexibility in the level of interactivity means drivers can adapt to appropriate levels of interaction based upon the demands of the driving task. Possible limitations of ISTs are associated with reductions of risk awareness and any positive effects tend to be very short-lived. Listening to the radio, although a potentially effective mitigation, may be more appropriate for in-car use as safety-critical in-cab audible warning alarms may be masked by the radio content. Similarly, mobile phone use does not seem a practical mitigation. Whilst there is some evidence that mobile phone conversations can improve driving performance, this is only the case in monotonous conditions. Furthermore, several other potential dangers are associated with mobile phone use, including driver distraction. Infrastructure and cab design could play a potential role in mitigating effects of low workload. Variable message signs have been shown to be effective in increasing arousal with several advantages.

2.3.4 Methods of predicting and measuring low workload

Dunn (2011) highlighted several methods for measuring workload. From a physiological perspective, it is assumed that workload and arousal are synonymous. Specific patterns of electrical brain activity, measured via electroencephalography (EEG), are known to indicate changes in arousal. These changes can in turn be used as indicators of workload. For example, slow waves (delta activity) and theta waves are thought to reflect low workload as they have been associated with drowsiness, sleeping and reductions in information processing capabilities (Grandjean, 1988; Lal & Craig, 2001). In contrast, a reduction in alpha and beta waves is typically observed during periods of low workload as these have been associated with waking or relaxed states and increases in arousal and alertness (Lal & Craig, 2001).

Heart rate (HR) and heart rate variability (HRV) can also provide a means of measuring workload (Lal & Craig, 2001). Reductions in HR have been linked to the onset of sleep states (low workload) and decreases in HRV have been linked to carrying out tasks that require some mental effort (higher workload) (Straussberger, 2006).

Other physiological measures include event related potentials (ERPs) which are extracted from the EEG signal and provide a measure of electrical brain activity following specific environmental stimuli. A decrease in amplitude of certain ERP components is associated with decreased arousal (Oken et al, 2006).

Eye movements (EOG) and blink rate have also been linked with changes in workload. Lack of eye movement or small and fast rhythmic blinks occur as drowsiness increases (Lal & Craig, 2001).

Other techniques to measure workload have been developed to take into account the subjective and experiential elements of workload. Hart and Staveland's (1988) NASA-Task Load Index (NASA-TLX) is the most common and accepted method for measuring workload (Dunn, 2011). The NASA-TLX differentiates between high and low self-reported workload dimensions. Three dimensions relate to task demands and three dimensions relate to subjective feelings of demand. The dimensions and descriptions are provided below.

Task demands

- Mental demand – the level of mental activity required to perform the task.
- Physical demand – the level of physical activity required to perform the task.
- Temporal demand – time pressure to perform task.

Subjective feelings of demand

- Performance – perceived success at performing task.
- Effort – the level of effort required to perform task.
- Frustration – the level of frustration or annoyance felt during performance of the task.

These individual scores are then combined to give an overall workload score.

May and Baldwin (2009) reviewed several technologies or techniques that are used to detect low workload in drivers. The technologies reviewed were generally designed to monitor psychophysiological signs of fatigue. The authors concluded from their review of the literature that video monitoring systems that record the extent to which the eyelid is closed over the pupil (PERCLOS systems) were one of the most effective methods for detecting changes in fatigue levels. Importantly, PERCLOS systems have also been shown to have a strong relationship with changes in driving performance (Dinges et al., 1998 & Mallis, 1999, cited in May and Baldwin, 2009).

Two other technologies reviewed by May and Baldwin (2009) included the S.A.M.G-3 steering attention monitor. This was described by Barton (2003, as cited in May & Baldwin, 2009) as monitoring equipment that records the corrective micro-adjustments made whilst a car driver is steering. If the micro-adjustments stop, this is assumed to indicate a lack of attention or task-related effort and an alarm can be sounded to make the driver aware their effort or attention has been reduced.

Head nodding assessment technology was also reviewed. These devices measure nodding of the head as this is a common behavioural symptom of drowsiness. May and Baldwin (2009) concluded that this technology was effective for detecting car driver drowsiness but it was prone to false alarms and it tended to detect drowsiness after driving performance had significantly reduced. As a result of this late detection it was not effective for preventing the negative effects of fatigue on driving performance.

There is some evidence that blink duration may be a valid measure of driver workload. Higher visual workloads in driving simulators have been associated with shorter blink durations and longer blink durations have been associated with increases in time-on-task (Benedetto, Pedrotti, Minin, Baccino, Re & Montanon, 2011). Increases in time on task can be associated with either underload or fatigue. There is some evidence that increased blink durations are associated with drowsiness and fatigue (e.g. Kobayashi et al. 1996); however, it is unclear whether blink duration would be an effective proxy measurement for observing low workload.

In addition to blink duration, blink rate may also provide another indicator of mental workload in operators. Zheng, Jiang, Tien, Meneghetti, Neely, Panton & Atkins (2012) found blink rates were effective for indicating the mental workload of surgeons during simulated surgical procedures. In general, the lower the blink rate the higher the workload.

RSSB published a research project that adopted and assessed several approaches to measuring driver workload (RSSB, 2004; T147). The project identified two tools for measuring driver workload: the train driver integrated workload scale tool (IWS) and the train driver DRAWS tool. The IWS tool can be used in 'real world' scenarios and gives feedback to the assessor in 'real' time. The IWS scale is based upon subjective ratings of mental workload that range from *not demanding* to *work too demanding.* Importantly, this tool could provide a useful indicator for low workload.

The DRAWS tool can assess the influence of task demands on the driver's workload. This tool provides a workload score on a 4-dimensional scale. The four dimensions are associated with information about the task, the mental operations of the task, responding to the task and time pressures of the task. The subjective ratings are given a score by the driver and can range from *no demand* to *maximum acceptable demand.* Importantly, this tool could be useful for distinguishing between low workload and acceptable workload.

2.4 Summary

In summary, workload is related to arousal. Low levels of workload are associated with sub-optimal performance on a task, narrowed attention or alertness on a task and less motivation to perform a task. Certain characteristics of a task have been associated with low workload. These include tasks that are monotonous, tasks that are performed for an extended period of time, and an objectively low level of work or difficulty associated with the task. Common effects of low workload on an operator include a general reduction in task performance, loss of motivation to perform a task, higher levels of boredom and fatigue. Several methods for mitigating the negative effects of low workload in driving situations have been researched with some success. These mitigations are generally based on the principle that adding extra cognitive load onto the driver under low workload situations will improve driving performance. Several methods have been used to measure workload. As workload is a hypothetical construct the most effective way to measure workload is to take multiple measures. Some of the most effective measures of workload that were reviewed included subjective measures such the NASA-TLX, physiological measures such as EEG, heart rate and heart rate variability, and behavioural measures such as reaction times and the amount the eyelid is closed over the pupil.

3 Review of train driving tasks

The train driving task was analysed in order to identify those activities which are least demanding and any situational characteristics that could have a bearing on task demand. The results of this analysis would provide a focal point for the remainder of the study, by identifying those train driving tasks and situations where the likelihood of low workload and associated errors is greatest. The aim was not to distinguish between low, optimal and high workload; rather, it was to identify train driving tasks that are relatively undemanding, for the remainder of the study to look into in more detail.

3.1 Method

This piece of work drew on task demand factors identified during the literature review (a detailed list can be found in Section 3.1.2) and used existing task analyses of train driving.

3.1.1 Analysis of the current GB train driving task

This work used unpublished task analyses produced as part of RSSB project T084, *The impact of ERTMS on driver workload* (HEL, 2004) to carry out a review of the current GB train driving task. The task analysis was very detailed, spanning over thirty-five A3 sheets of paper.

For the purposes of the work reported here, the analysis of the current GB train driving task was simplified and assigned ratings in order to identify the presence or absence of task demand factors (as described on Sections 3.1.2 and 3.1.3). Only the first four levels of the task breakdown were assigned ratings, although in doing so the nature of the tasks in the 5th and 6th levels of the breakdown in the original task analysis were taken into account. This meant that ratings were assigned to just over 200 train driving tasks.

Since 2004 there have been some changes to the operational rules governing activities on the railways in Great Britain, so an operations specialist was consulted to ensure that where significant changes have occurred, the analysis took the new rules and associated task demands into account.

3.1.2 Task demand factors

The literature (Chapter 2) did not identify a method for predicting task demand based on task descriptions. It did, however, identify several task characteristics that have the potential to lead to underload. The characteristics below are a synthesis of those identified by Hart and Staveland (1988) and MacDonald (2003). These characteristics were treated as factors increasing the likelihood of low demand, when reviewing the train driving task:

- Sensory/perceptual –
    - difficult to see or hear stimuli,
    - rare stimuli (spatially or temporally),
    - repetitive environmental stimuli,
    - predictable environmental stimuli,

- compatibility between form of stimulus and the information it is intended to convey,
- compatibility between stimulus information and individual's information processing abilities.

- Central processing –
  - easy or simple decision-making,
  - low decision rates,
  - low information processing rates,
  - low level of attention required,
  - low requirement for long-term memory,
  - low requirement for concentration,
  - no distractions,
  - low awareness of consequences of performance error.
- Psychomotor –
  - high response certainty,
  - no possible action alternatives,
  - low required response rate,
  - high degree of timing certainty,
  - low requirement for accuracy,
  - high compatibility between stimulus and required action.
- Physical –
  - low requirements for large physical effort (e.g. lifting),
  - low requirement for fine movements (e.g. gripping),
  - low requirement to move or change posture,
  - high levels of physical comfort,
  - unable to walk around.
- Affective –
  - low need for social engagement,
  - no emotional reactions or stress
- Time-on-task: Effects generally noticeable after 15-30 minutes but under certain conditions time-on-task effects can be seen much earlier.

These factors combine in complex ways that are not yet understood in order to influence workload and perceptions of workload. For example, the presence of demanding factors can compensate for undemanding factors within a task, to a degree. The method used to carry out this piece of research does not take into account these complex interactions and so does not distinguish between low, acceptable and high workload. Instead, it highlights those tasks which have fewer demanding factors and more factors associated with low demand.

3.1.3 Coding

A spreadsheet was developed, an extract from which is shown in Table 1. Each row represented a train driving task, and each column represented a task demand factor such as 'low decision rates' or 'high response certainty'. The cells in the spreadsheet were used to express whether the low demand factor was present or absent in the task (i.e. does the task of 'Acknowledge train manager initiated codes' have a high response certainty? A red-amber-green rating was assigned in order to express whether the low demand factor was present in the task. If a task did have a low demand factor it was given a red rating and if it did not have that factor then it was given a green rating. For example, the task 'Acknowledge Train Manager initiated bell codes' during train dispatch requires the driver to press a button one or more times. This task was given a red rating to indicate that it **did have** a low requirement for large physical activity such as lifting, but a green rating to indicate it **did not have** a low requirement for small movements. This may seem counterintuitive, but because the presence of a low demand factor increases the likelihood of low workload, tasks with the more red codes were of more interest to the study.

Amber ratings were assigned if the factor may or may not be present depending on situational variables such as the diagram or rolling stock. Where amber ratings were used, the potential situations in which tasks could have low demand were recorded.

Table 1: Design of spreadsheet used for coding

	Factor 1	Factor 2	Factor 3
Task 1		If the task lasts more than 15 minutes	
Task 2	If there is less than 2 minutes between stations		If driving a freight train
Task 3			

Once the analysis was completed the coding was used to determine which tasks (and under what circumstances) place a low demand on the driver. The number of red, amber and green ratings was counted for each task, and a task was identified as being potentially undemanding if:

- It had a high number of red ratings, and
- It had a low number of green ratings.

A task was said to have a high number of red ratings if it was in the top 25%, when all tasks were ranked by the total number of red ratings for the task. It was said to have a low number of green ratings if it fell within the bottom 25% of tasks, when ranked by total number of green ratings.

3.1.4 Understanding the effects of ERTMS on workload

The study considered the implications of Level 2 European Railway Traffic Management System (ERTMS) on train driver workload through the analysis of existing research in the area. The design of the in-cab interface for ERTMS, driver tasks and the procedures surrounding its use have evolved in recent years, giving rise to the possibility of previous research being out of date. Thus, discussions were also held with RSSB's Human Factors Specialist working on the ERTMS driver-machine interface (DMI), and ATOC's ERTMS Operations Specialist, who is also an ex-train driver and has driven trains with and without ERTMS.

The project team considered Level 2 ERTMS to be the implementation most relevant to the GB context. This involves signalling information being provided in-cab only (apart from some trackside indicators). This is the system currently being trialled on the Cambrian Line prior to wider deployment across Great Britain.

3.2 Outcome of the review of the current GB train driving task

3.2.1 Tasks that place lower demands on the driver

The following train driving tasks were highlighted as potentially placing lower demands on the driver. These are tasks which were in the top 25% of tasks, when all tasks were ranked by the total number of red ratings for the task, and within the bottom 25% of tasks, when ranked by total number of green ratings. For the current analysis this was equivalent to tasks where more than ten of the twenty-eight low demand factors were rated as red (low demand), and fewer than seven were rated as green (higher workload).

This does not mean that the demands of these tasks are so low that drivers are likely to make errors, nor does it mean that underload related errors are unlikely in other train driving tasks. The list below simply identifies tasks that this study should look at in more detail, because they have characteristics associated with low demand.

- Evaluating the speed that the train should be driven at. This primarily involves taking in information from the environment and retrieving information from memory to determine what the appropriate speed is. It takes place throughout the drive, in parallel with the next task of controlling the train speed. It is heavily based on training, including route and traction knowledge, but is reinforced by trackside signs. On routes that drivers are highly familiar with and where signal aspects are highly predictable, speed strategy would also be highly predictable.
- Driving according to the speed strategy. This is a monitoring and control task whereby the driver of the train makes speed adjustments and checks the outcome of those changes, comparing them with the speed strategy. On routes that drivers are highly familiar with and where signal aspects are highly predictable, speed control actions and their outcomes would be highly predictable.
- Checking the authority to proceed. This task also takes place throughout the drive, and involves checking signals. The signal aspect tells the driver if the train is approaching a red signal and needs to slow down or stop. The nature of the operation of the rail network in Great Britain is such that the aspect displayed by some signals is highly predictable, whereas for others it is not.

- Ensuring the train is safely stopped (i.e. not going to roll forwards or backwards), when the train has arrived at a station. This is likely to be a very easy and highly practiced task, but of very short duration.
- Checking platform end signal for the Close Doors Indicator during platform staff assisted dispatch. During this task the driver of the train does not need to engage with what is happening on the platform, but simply has to wait, notice when the Close Doors indicator by the signal illuminates, and then close the doors. The length of the monitoring task depends on the dwell time.
- Driver-guard station duties. The driver does not need to engage with the dispatch task and simply has to wait for the guard to ring the bell/buzzer system to indicate that the train can depart the station (as long as the signal is clear). The length of time the driver waits depends on the dwell time.

Tasks to do with signing in for duty, preparing trains and most of the tasks associated with handing over or securing trains at the end of the shift did not fit the criteria used during this analysis (Section 3.1.3) and were not considered for further study.

3.2.2 Key situational factors that can influence demand

The analysis also identified many situational factors that would affect workload. These resulted in an amber rating and were considered as being important factors in the design of the trial (see Chapter 5). These were:

- Design and layout of information, such as schedule cards, operating notices. The more difficult they are to read the higher the demand of the task.
- The diagram, and in particular:
    - Repetition: frequent stops, short routes worked back and forth, high number of shunt moves.
    - Predictability: unvarying signal spacing, diagram, traffic patterns and signal aspects received.
    - Length of time spent doing a task such as driving between stations or signals, time spent at stations.
    - Absence of requirement for response (e.g. lack of changes to speed restrictions, lack of restrictive signal aspects).
    - Absence of distractions.
- Familiarity with the diagram, location or rolling stock.
- The driver's stress levels and emotional response to the situation (although in most cases it is likely that the driver will be calm and not under stress).
- The environment, including:
    - How repetitive environmental features are.
    - How predictable environmental features are.
- Rolling stock/cab design, for example:
    - How comfortable it is.
    - Whether drivers can vary their posture (i.e. drive standing or sitting).

- Type of operation
  - Driver-Only Operation (DOO), where demand during some tasks such as dispatch is higher than when there is a guard or platform staff to assist with tasks.
  - Driver and guard operation, where some tasks are split between the driver and guard, resulting in lower demand on the driver during some tasks such as dispatch.
  - Platform staff assistance at stations, with demand being lower than DOO during the dispatch task.
- Communication
  - The complexity of the message being conveyed or received, with increasing complexity increasing demand.
  - Predictability of the message being conveyed, with increasing predictability reducing demand.

3.2.3 Implications for the simulation trials

This review of train driving tasks identified that driving the train between signals and stations on a simple and familiar route with little or no action required to control the train (i.e. no need to accelerate or decelerate significantly, and also waiting to depart stations when platform duties are carried out by platform staff or a guard), had characteristics associated with low demand. These outcomes are considered further in section 4.

3.3 Implications of ERTMS on train driver workload

The design of the in-cab interface for ERTMS, the procedures surrounding its use and the resulting driver tasks have evolved in recent years. Level 2 is in use on one railway line in Wales, and this has provided an insight into the train driving behaviours associated with the use of ERTMS in GB. Nevertheless, there are still questions over how ERTMS impacts upon workload in various train driving situations, and it is not clear whether it would increase or decrease the likelihood of underload. This section draws on three relevant studies in this area:

- A comparative cognitive task analysis of the different forms of driving in the UK rail system (Buksh et al., 2013)
- ERTMS Train Driving InCab vs. Outside - An explorative eye-tracking field study, by (Naghiyev et al., 2014)
- T084, Impact of the European Rail Traffic Management System on driver workload (Human Engineering Limited, 2004 on behalf of RSSB).

None of these studies claim to provide a definitive and comprehensive assessment of the potential for low workload with ERTMS; however, they have begun to identify where there exists a potential for a change in workload, at a macro level.

Naghiyev (2014) highlights that a key feature of ERTMS in terms of driver behaviour is the integration of movement authority and speed information, in the form of a speed/braking curve presented on the driver-machine interface (DMI). Currently drivers assimilate information from a wide range of sources such as their route knowledge, the current status of the track, signal aspects and speed restrictions in order to determine a suitable speed of travel. Thus, at a macro level, and during normal operation, the cognitive load associated with speed choice would appear to be lower with ERTMS than the current train driving task (Buksh et al., 201). This overall conclusion is supported by the analysis carried out by HEL (2004). They state that ERTMS where all signalling indications including movement authority and routing are displayed in-cab (i.e. Level 2 ERTMS without signal overlay) results in lower workload than the current train driving task. HEL (2004) conclude that Level 2 ERTMS involves:

- Fewer psychomotor tasks.
- Fewer visual tasks (as information is combined on the in-cab display).
- Up to 1000 fewer cognitive tasks.
- Reduced time pressure to detect information (because it is displayed in-cab rather than at a fixed point outside the cab).

This suggests that Level 2 ERTMS without lineside signals could reduce workload for train drivers. However, studies have identified a range of factors that could have an effect on workload under ERTMS:

1. **ERTMS implementation:** Where lineside signals are overlaid, the monitoring of both the DMI and lineside signals increases workload. HEL (2004) conclude that where lineside signals are overlaid, this increases workload relative to the current GB train driving task.
2. **Driver strategy**: Buksh (2013) states that ERTMS would enable drivers to become reactive, because the system determines the appropriate speed for the location and available movement authority and uses auditory alerts to warn the driver of required changes. Indeed, it intervenes if the driver fails to react. Naghiyev (2014) found that drivers spent significantly longer looking at their speedometer when driving under ERTMS, rather than monitoring the environment. They attributed this, in part, to a shift from a proactive, anticipatory driving strategy, towards a more reactive, monitoring strategy. This reactive strategy may be associated with reduced workload. (Buksh et al., 2013; Naghiyev, 2014). There is a clear interaction between driver strategy and motivation, which the literature review (see Chapter 2) has highlighted as being an important factor influencing feelings of boredom and thus arousal.
3. **Type of traction:** The acceleration and braking capabilities of trains anecdotally affect how reactive drivers can become. The braking technique for high performance modern rolling stock may enable the driver to simply respond to the indication given on the in-cab display, but for freight trains, and Class 43 stock, drivers would need to follow a more complex braking technique in order to achieve the changes in speed required by the signalling system. Thus, they are unlikely to be able to take a reactive approach to driving and their workload is likely to be higher than drivers of high-performance rolling stock.
4. **When ERTMS is first introduced** cognitive load is likely to be higher where drivers' previous route knowledge (e.g. in relation to braking points) conflicts with the braking profiles programmed into ERTMS (Buksh et al., 2013).

5. **Transitions into and out of ERTMS** may be associated with increased workload (HEL, 2004)

On the whole, none of the studies carried out so far have been able to investigate fully whether ERTMS makes it any more or less likely for the driver to experience cognitive underload. However, it is important to note that ERTMS has a supervisory component which (when enabled) alerts drivers when the required control inputs are not made, and intervenes if the required input is not made following the alert. Thus, many of the potential consequences of low workload are mitigated.

4 Identification and selection of mitigation

Having identified train driving tasks which have characteristics associated with low demand, the study then focused on identifying a mitigation which could be applied during a simulator trial.

The literature review highlighted that underload is a product of task and environment characteristics and the individual's response to the task. Therefore, measures that affect the individual's response to the task were considered alongside those that modified what the person was asked to do. In addition, measures that reduce the impact of underload-related errors were also considered.

The literature review identified a small number of potential measures (see Chapter 2). A half-day workshop was held to:

- Decide which train driver task would be studied during the simulation exercise (because the choice of task would influence the measures selected to mitigate low demand).
- Brainstorm and compile a list of potential measures that might mitigate low task demand or its negative impact on driver performance.
- Select one or two promising and implementable measures, with a view to selecting a single measure to trial, once the constraints of the trial set up and environment were known.

The workshop was attended by six members of the project team from RSSB and TRL. It began with a description of the train driving task and particularly those tasks which were identified previously as placing low demand on the driver.

The workshop attendees agreed that the remainder of the research would focus on a long drive without station stops, and with little requirement for control inputs. This was because:

- Previous analysis highlighted associated tasks to be potentially undemanding (see Chapter 3).
- Task duration exacerbates the likelihood of errors due to low demand, and the duration of this task can be significantly longer than half an hour.
- This task is particularly suited to being studied in a train driving simulator (whereas dispatch, for example, is not).

The mitigations identified during the literature review were presented and discussed and then a brainstorming activity was completed to identify other potential measures. Workshop attendees discussed the evaluation criteria for the mitigations before the brainstorming activity, so that they could consider different ways of mitigating the problem. The evaluation criteria included:

- Addresses task/environment factors associated with low workload (e.g. repetition).
- Improves the individual's response to the task (e.g. motivation).
- Increases the system's resilience (make underload related errors less of an issue).

- Likely effectiveness of the measure.
- Potential negative side effects.
- Implementability (for trial purposes only).

It was decided that the purpose of the simulation study was to demonstrate that low workload can be addressed rather than to recommend a particular mitigation. This decision meant that more innovative ideas could be considered, even if there may be difficulties implementing the measure in the real world environment with current technology.

A list of measures was produced for evaluation. Pre-screening of the measures resulted in some of them being eliminated from the list. All attendees then took part in the evaluation of the list of mitigations. Each attendee was asked to evaluate each of the mitigations individually based on the criteria listed above. They were then asked to pick and name the top three measures based on their evaluation. A group discussion followed in which a refined short list was produced. Each group member was then asked to choose their top two measures, and a further discussion took place to decide which two would be carried forward into the next stage. Although only one measure was going to be used in the simulation exercise, at this stage there was uncertainty as to which facility the experiment would be conducted at, and about the feasibility of implementing some of the ideas generated during the workshop.

The full list of measures that was discussed is shown in Appendix A. The two selected for further consideration were:

1. An interactive secondary task
2. A combination of the following measures:
    a. briefing about low workload,
    b. a workload rating scale exercise administered while driving (to increase workload and also make drivers aware of their workload levels),
    c. commentary driving, applied when task demands are low.

The first of these two measures made it into all workshop participants' final list and was thought to be promising.

The latter measure was considered relatively easy to implement, however, concerns were raised that (in the context of a simulation trial) it would simply amount to telling people to try harder during one of the drives.

Because of these concerns, playing music in the train cab was given further consideration as a replacement. Specifically, music developed by Brodsky & Kizner (2011) was considered to be promising as it had been designed specifically to "…furnish a driving environment that maintains alertness and positive mood without diverting cognitive resources" (Brodsky & Kizner, 2011 p. 10).

It was decided that an interactive secondary task and music would be taken forward to the pilot trials, at which point the interactive secondary task received better feedback from the pilot participant, and the music measure appeared to generate a degree of annoyance. As a result the interactive secondary task was selected for use during the trials.

5 Simulation trials

5.1 Method

5.1.1 Participants

Ten participants were recruited to take part in the experiment. Of these, one did not complete the second drive of the trial, so data was analysed for nine participants. All participants were male and employed by various train operators.

Participants' ages ranged from 44-61 years with an average age of 51.4. The number of years train driving experience for the participants ranged from 5-17 years with an average experience of 12.3 years.

Six of the participants were current drivers, and the remaining participants were a driver trainer, competence manager and competence developer. Although there were variations in job role, all participants were competent train drivers with their assessments up to date.

5.1.2 Experimental design

The study was designed to test the effectiveness of a mitigation technique to support driver alertness and attention in situations of potential cognitive underload. It employed a repeated measures design with participants required to complete two drives on Southern's train simulator of the same simulated route. The two drives were held over two separate sessions on different days. One of the drives was conducted without the mitigation and provided a baseline measure of driver performance. The other drive incorporated the mitigation in the form of a secondary verbal task consisting of task relevant questions that were administered by the experimenters. The sessions were counterbalanced so that half of the participants completed the baseline drive first, and half completed the mitigation drive first.

Participants were asked to complete the two drives on different days so as to minimise practice effects and cumulative fatigue from influencing their performance on the second drive. The two drives were also carried out at similar times of the day in order to control for possible confounds that may be caused by a participant's circadian rhythm.

A direct evaluation of the effectiveness of the mitigation was carried out via a range of performance related measures, participants' subjective assessments of both drives, and physiological indicators of mental workload.

In summary, the following outcome measures were assessed:

- Driver performance measures
    - Speed choice (mean, frequency of over and under-speed events, variability of speed).
    - Detection (number of responses) and reaction time to events that would typically be experienced in an operational environment, in the form of the presence of track workers.
    - The detection (number of responses) to unexpected events that would not typically be experienced in an operational environment, in the form of Automatic Warning System (AWS) irregularities.

- Detection and reaction time to a critical safety event in the form of level crossing barriers being open (presented near the end of the participant's second drive).
  - Frequency of driver safety device (DSD) pedal activations.
- Subjective measures
  - Perceived workload
  - Subjective stress and arousal levels
  - Sleepiness/alertness ratings
  - Acceptance of and attitude towards the verbal secondary task as a possible mitigation for cognitive underload (as assessed in the interview process).
- Physiological measures
  - Heart rate as an indicator of mental workload.

5.1.2.1 Independent variable

The single independent variable that was manipulated was the presence or absence of the mitigation.

5.1.3 Materials

5.1.3.1 Simulator

Testing was conducted using Southern's CORYS train simulator which is located at the Traction and Rolling Stock Maintenance Depot in Selhurst, South London.

The simulator consisted of a realistic train cabin environment which replicated a Southern Class 377 locomotive. The controls included a throttle/brake controller and all other switchgear and visual feedback displays associated with this type of train. All of the control interfaces had a realistic feel and could be used in the normal manner. The cabin also included fully functioning automatic warning system (AWS) visual and audible indicators, as well as a driver vigilance monitor in the form of the "driver safety device" (DSD) pedal.

The virtual environment was projected onto a screen in front of the driver, while a stereo sound system with speakers generated realistic train and warning sounds which further helped to replicate a realistic driving experience.

Figure 1: Southern train simulator displaying the simulated route used in this study

5.1.3.2 Route

A virtual route consisting of a 23.7 mile stretch of track was used. In each drive, participants were required to drive the route in both directions, resulting in a total route length of 47.4 miles. The direction in which drivers completed the route was fixed for all participants.

The virtual environment consisted of a mostly rural landscape with areas of open green fields interspersed with sections of trees, and a smaller number of sections with buildings and other basic trackside furniture. The track layout contained a mixture of straight sections and curves, as well as several level crossings, tunnels and bridges. The route also passed through 13 stations, with the 13^{th} station acting as the finishing position for the outward drive, and the starting position of the return drive.

In order to further enhance the monotonous nature of the route, participants were informed that a blanket speed limit of 30mph was in force along the entire route. This resulted in an approximate drive length of 2 hours per drive.

5.1.3.3 Events

In order to obtain appropriate measures of driver attention and situation awareness, participants encountered several events during the drives to which they were required to respond. Driver responses to the events allowed for a direct comparison to be made between the baseline drive and the drive with the mitigation.

The events, and the responses required by the drivers, were categorised as follows:

- An event which a driver *would typically expect* to encounter whilst driving in a real operational environment.

- These events involved the presence of **track workers** in high visibility clothing who were positioned at three separate locations along the route. The track workers were positioned in such a way that drivers needed to respond as quickly as possible once they became visible (i.e. they were located on curves or partially obscured by trees etc.). Moreover, the track workers were positioned such that as soon as they became visible, the driver could determine that they were at a safe distance from the running line, and were not in the path of the train.
- Consistent with standard operational procedures, participants should sound the horn as soon as they detected the presence of a track worker.

- An event which a driver *would not typically expect* to encounter whilst driving in a real operation environment
    - Four **automatic warning system (AWS) irregularities** were presented at appropriate intervals along the route. The irregularities were the absence of an audible tone when the driver passed over the AWS magnet. The AWS irregularities were placed at suitable intervals so as to maintain a reasonable level of unpredictability.
    - In response to the detection of AWS irregularities, participants were required to press and release the hill-start button (located on the end of the brake/throttle controller) as quickly as possible. Although this is not how drivers would normally respond to such an event in an operational environment, it allowed the participants to provide responses quickly without having to take their hand off the controller or take their eyes off the virtual environment.
    - For events such as these, drivers would normally report to the signaller. But for the purpose of the study and to avoid any social interaction the hill-start button was deemed adequate to help induce monotony.

- Critical safety event
    - Towards the end of their second drive, participants encountered a critical safety event in the form of the train being signalled over a **level crossing with barriers remaining in the "open" position**.
    - In the same way as for the AWS irregularities, participants were required to press and release the hill-start button as quickly as possible in response to the level crossing event.
    - For the same reason as for the AWS irregularities the hill-start button was chosen to avoid social interaction with a signaller.
    - As this critical event was likely to be highly memorable to the participants, it was only presented in the second drive. The level crossing event was therefore included in half of the baseline drives and half of the mitigation drives.

A summary of all event locations and the direction in which the events are encountered (outward vs. return) are listed in table 2. The events were in the same location during the baseline and mitigation drives, giving two identical scenarios and allowing experimental control over any increase in alertness caused by the trackworkers and AWS irregularities. The exception to this was the level crossing failure, which only featured in the participant's second drive. Drive order was counterbalanced to eliminate order effects.

Table 2: Event locations and driver responses required

Event	Location (distance from station 1 in miles)	Drive direction	Driver response
Track worker 1	10.17	Outward	Sound horn
AWS irregularity 1	14.98	Outward	Press hill-start
AWS irregularity 2	17.18	Outward	Press hill-start
AWS irregularity 3	21.16	Outward	Press hill-start
Track worker 2	20.23	Return	Sound horn
AWS irregularity 4	15.15	Return	Press hill-start
Track worker 3	8.93	Return	Sound horn
Level crossing barrier	4.83	Return	Press hill-start

5.1.3.4 Mitigation – verbal secondary task

The verbal secondary task consisted of a series of questions that the experimenter asked the participant over the intercom during the mitigation drive condition. The questions were task relevant, requiring the driver to attend to and develop an awareness of the track environment, read the instruments and controls in the cab and think about possible emergency scenarios. In order to minimise potential distraction, the questions did not require long or complex answers and the drivers were informed that they did not have to answer if they were encountering a difficult driving situation or felt that the question was a distraction at that particular moment.

The questions were asked at fixed locations spaced one mile apart during the mitigation drive, resulting in a total of 45 questions being asked. A full list of the questions used is presented in Appendix B.

5.1.3.5 Questionnaires and interviews

In order to obtain information relating to participant demographics, subjective ratings of driver performance, fatigue levels and driver impressions of the mitigation, several questionnaires and a post-mitigation drive interview were administered. These are described below.

1. Demographic questionnaire (see Appendix C) – To collect basic demographic information about the participant (e.g. age, gender, years of train driving

experience etc.). This questionnaire was administered before the participant commenced their first drive

2. Karolinska sleepiness scale (KSS) (see Appendix D) - To gain an understanding of participant's level of sleepiness/alertness. Drivers rated their level of alertness on a scale of 1 (extremely alert) to 9 (extremely sleepy). Drivers were asked to give a rating three times; immediately prior to commencing the first drive, during the brief pause half way through each drive after participants stopped at station 13, and at the end of each drive.
3. NASA TLX (see Appendix E) – To collect a subjective measure of mental workload. The NASA TLX was administered immediately after each drive.
4. Stress Arousal Checklist (SACL) (see Appendix F) – To investigate how completion of each drive affected each participant's stress and arousal levels. The SACL was administered immediately after each drive.
5. Post mitigation interview – To obtain subjective impressions of the verbal secondary task as a technique for mitigating the effects of cognitive underload (see Appendix G). The relevant post drive interview (A, B or C) was conducted after the participant's drive.

5.1.3.6 Physiological recording equipment

A Polar RS800 heart rate monitor, which consists of a band worn around the chest in contact with the skin and a wrist watch, was used to record participant's heart rate. Participants were asked to wear the heart rate monitor during both drives, although its use was not compulsory.

5.1.4 Procedure

Upon arrival, participants were provided with some basic information about the trial and were asked to provide informed consent prior to taking part.

5.1.4.1 Trial schedule

The test procedure for each of the drives followed the schedule outlined in Table 3.

Table 3: Participant schedule

Trial start +:	Duration	Item
00:00	00:05	Arrive, welcome
00:05	00:05	Familiarisation drive
00:10	00:10	Setup, fitting of physiological monitoring equipment
00:20	00:05	Pre-drive questionnaires
00:25	02:00	Simulated train driving task
02:25	00:15	Post-drive questionnaires
02:40	00:05	Removal of physiological monitoring equipment
02:45	00:00	Session end

5.1.4.2 Familiarisation drive

Prior to the start of each experimental drive, participants were asked to drive a short route in order to familiarise themselves with the controls of the simulator and the experience of driving in a virtual environment. Participants drove as they normally would for approximately 5 minutes. This also provided participants with an opportunity to ask any questions before the start of the main trial, and allowed them to make any necessary adjustments to the cab so that they were comfortable prior to commencing.

5.1.4.3 Experimental drives

A standard set of instructions were provided to the participants prior to commencing the drives. They were asked to drive as they normally do and were reassured that their driving was not being judged.

Participants were instructed to drive a non-stop service from station 1 to station 13 without stopping at any of the stations. Participants were informed that once they reached station 13, there would be a brief pause while another scenario was being loaded, after which they would then be required to drive a non-stop service from station 13 back to station 1.

Prior to commencing their drives, participants were provided with a short briefing to explain the effects of workload on task performance, and that performance typically decreases when arousal levels are either too high or too low. To help facilitate this briefing, participants were also shown the inverted U-shape curve showing how performance increases and then decreases again with higher levels of arousal. The purpose of this briefing was to reassure participants that the verbal secondary task was not intended to distract them from the primary driving task, whilst at the same time not making it obvious to the participants that the questions were designed specifically to improve their performance (which could have artificially influence how they approached the driving tasks).

Participants were informed that a blanket 30mph speed restriction was in force along the entire route due to extreme high temperatures, and that they should drive as they normally would if such conditions were presented to them in a real operational environment. They were also told that if they encountered any events for which they would normally stop and report to the signaller, they should press and release the hill-start button as quickly as possible. Participants were reassured that, although pressing the hill-start button may have felt like an unnatural response, it was simply being used in the experiment as a means for collecting their responses. Participants were not told exactly what events they would encounter so as not to prime them.

Immediately before commencing the drives, the experimenter informed the participants that everything had been set up for them and that they could assume that all of the pre-start checks and routines had been carried out. They were then informed that they had received the "right away" and were free to set off when ready.

5.1.5 Hypotheses

5.1.5.1 Hypotheses relating to driver performance and attention

H01 There will be a reduction in the frequency of under and over-speed events (+/- 3mph from the imposed 30mph speed limit) in the mitigation drive compared with the baseline drive.

H02 There will be a reduction in speed variability in the mitigation drive compared with the baseline drive.

H03 Mean speed in the baseline drive will differ when compared with the mitigation drive.

H04 Participants will respond more quickly to the track worker events in the mitigation drive compared with the baseline drive.

H05 Participants will detect a greater number of track workers in the mitigation drive compared with the baseline drive.

H06 Participants will detect a greater number of Automatic Warning System (AWS) irregularities in the mitigation drive compared with the baseline drive.

H07 Participants will respond more quickly to the level crossing critical safety event in the mitigation drive compared with the baseline drive.

H08 Participants will perform fewer driver safety device (DSD) activations in the mitigation drive compared with the baseline drive.

5.1.5.2 Hypotheses relating to subjective assessment of driver performance

H09 After the mitigation drive, participants will report higher mental workload on the NASA-TLX compared with the baseline drive.

H10 Participant's perceived levels of fatigue will be lower in the mitigation drive compared with the baseline drive

H11 After the mitigation drive, participants will report higher arousal levels on the SACL compared with the baseline drive.

5.1.5.3 Hypotheses relating to physiological indicators of mental workload

H12 Participant's heart rate will be increased in the mitigation drive compared with the baseline drive.

6 Results

6.1 Performance measures

6.1.1 Over and underspeed events

Hypothesis H01 was that there would be a reduction in the frequency of under and over-speed events (+/- 3mph from the imposed 30mph speed limit) in the mitigation drive compared with the baseline drive.

Driving speeds were sampled at each of the 162 AWS magnets that the train passed over, giving one measurement every quarter of a mile, on average.

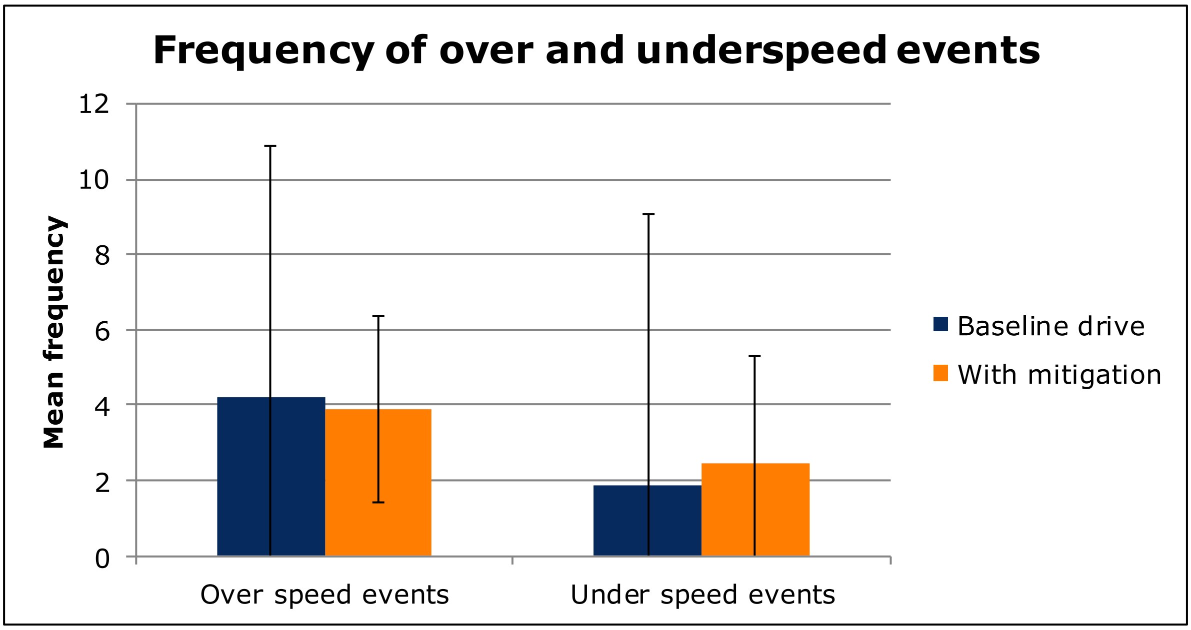


Figure 2: Frequency of over and underspeed events

Figure 2 shows that there was very little difference between the overspeed frequencies in the baseline and mitigation drives. A Wilcoxon Signed-ranks test found no significant difference in the number of overspeed events (Z=-.378, p=.705).

There were also very little difference between the two conditions in relation to the number of underspeed events, and again using a Wilcoxon Signed-ranks test, no statistically significant difference was found (Z=-.632, p=.527).

6.1.2 Speed variability

Hypothesis H02 was that there would be a reduction in speed variability in the mitigation drive compared with the baseline drive.

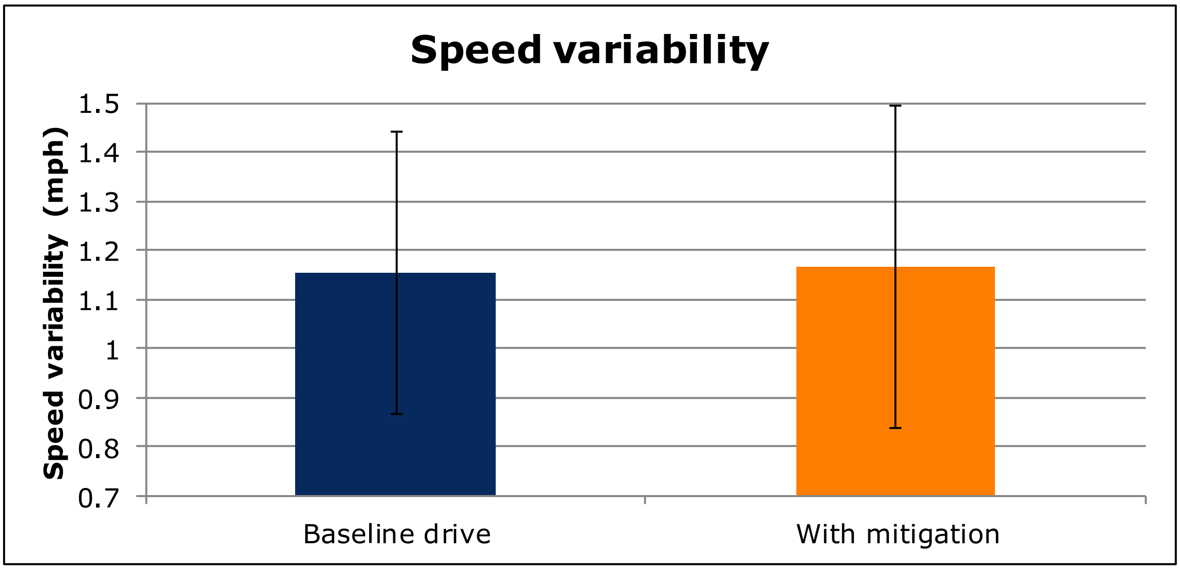


Figure 3: Speed variability

Figure 3 illustrates that there was very little difference in speed variability between the two conditions. A paired samples t-test was carried out and revealed that the difference between the two conditions in terms of speed variability was not statistically significant ($t(8) = -.189$, $p=.855$).

6.1.3 Mean speed

Hypothesis H03 was that mean speed would be lower in the baseline drive compared with the mitigation drive.

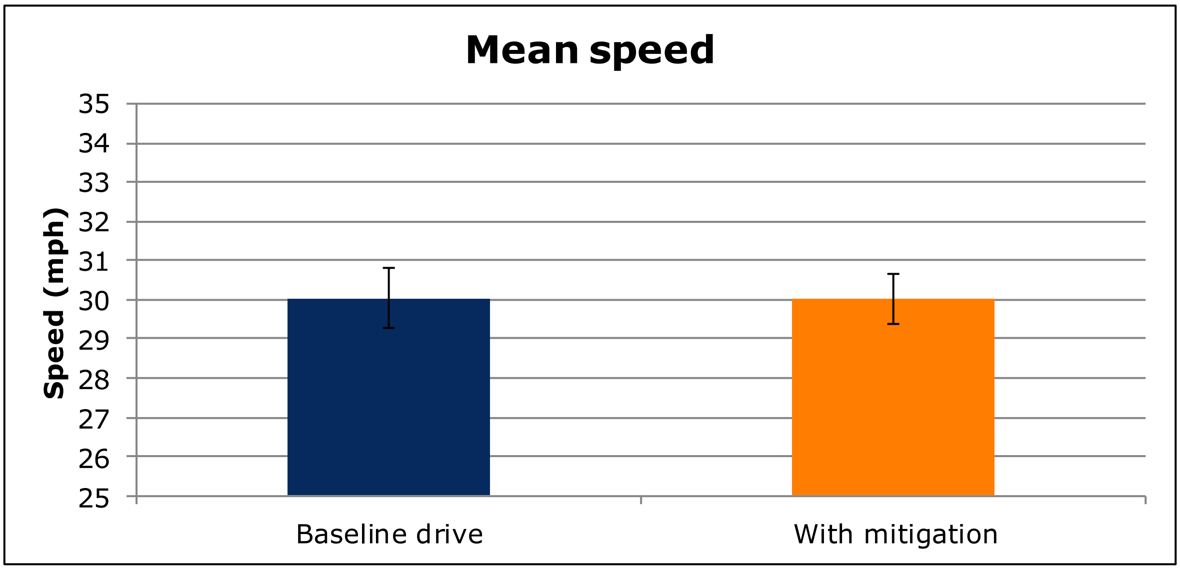


Figure 4: Mean speed

Figure 4 shows that the average speeds of the nine participants did not vary greatly between the baseline drive and the mitigation drive. A paired samples t-test showed that the mean speeds were not significantly different ($t(8) = .154$, $p=0.881$).

6.1.4 Track worker reaction times

Hypothesis H04 was that participants would respond more quickly to the track worker events in the mitigation drive, when compared with the baseline drive.

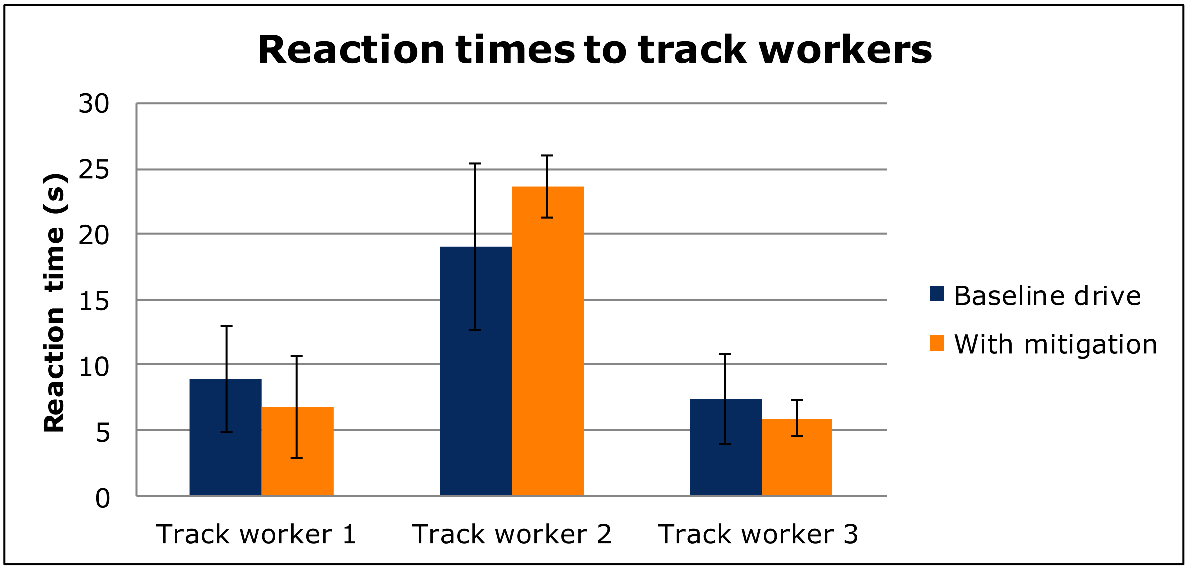


Figure 5: Reaction times to track workers

Figure 5 shows that there was no clear effect of the mitigation on reaction time to track workers, as participants seemed to react a little faster in the mitigation condition to track workers 1 and 3, but slightly slower in the mitigation drive to track worker 2. Paired sample t-tests show that the differences were not statistically significantly for track worker 1 ($t(8) = 1.48$, $p=0.177$), track worker 2 ($t(8) = -1.845$, $p=0.124$) or track worker 3 ($t(8) = 1.013$, $p=0.341$).

6.1.5 Response to track workers

Hypothesis H05 was that participants would detect a greater number of track workers in the mitigation drive, when compared with the baseline drive.

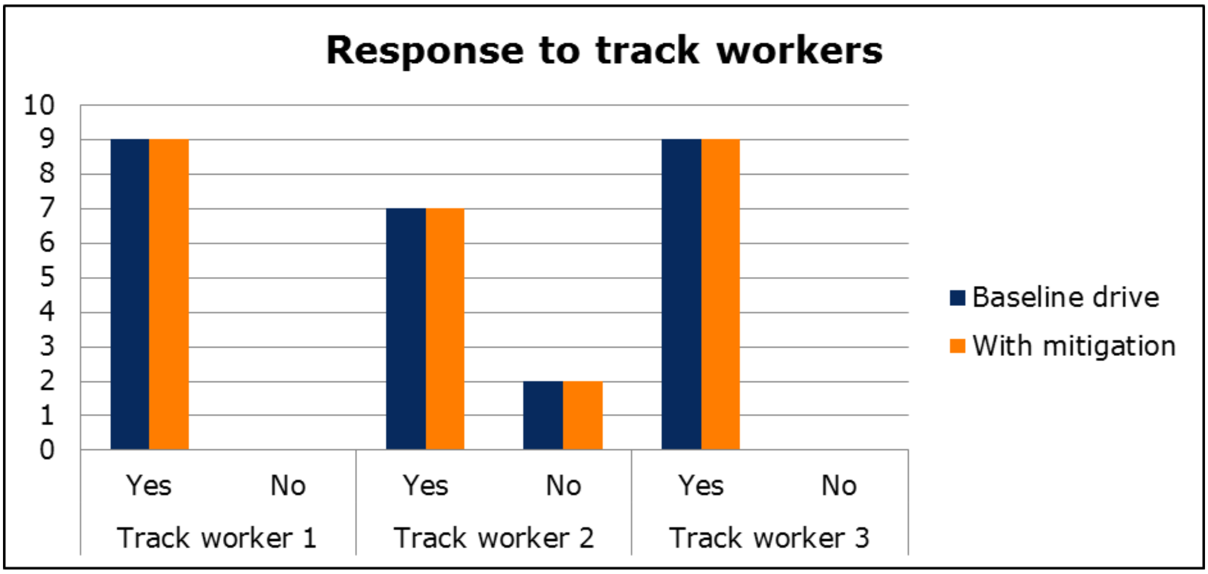


Figure 6: Response to track workers

Figure 6 shows whether participants responded to the three track workers. Both track worker 1 and 3 were identified on every drive. Track worker 2 was missed twice in both conditions. A chi-square test showed that there was not a statistically significant difference in response rates to track worker 2 (χ^2 (1, N=9) = 1.148, p=0.417).

6.1.6 *Response to AWS irregularities*

Hypothesis H06 was that participants would detect a greater number of Automatic Warning System (AWS) irregularities in the mitigation drive compared with the baseline drive.

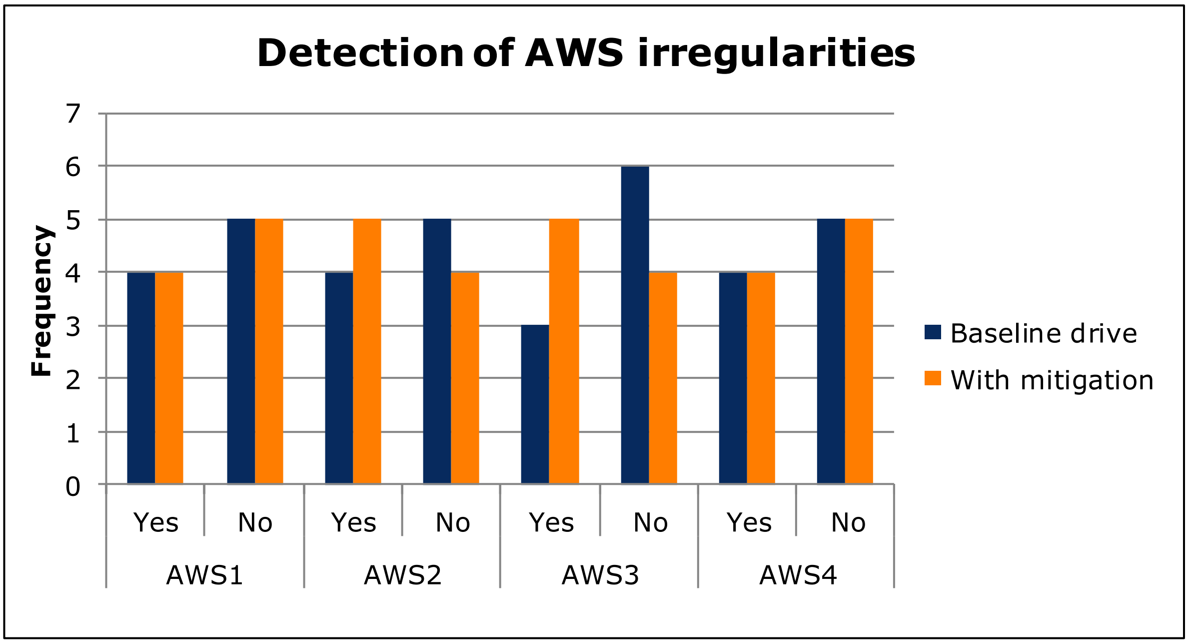


Figure 7: Response to AWS irregularities

Figure 7 shows that a large number of AWS irregularities were missed in both drives and that he difference between conditions was minimal. Chi-squared tests confirmed that the differences were not statistically significant for AWS1 (χ_2 (1, N=9) = .090, p=.643), AWS2 (χ_2 (1, N=9) = .090, p=0.643), AWS3 (χ_2 (1, N=9) = .225, p=.595) or AWS4 (χ_2 (1, N=9) = 2.723, p=.167).

6.1.7 *Reaction time to level crossing failure*

Hypothesis H07 was that participants would respond more quickly to the level crossing critical safety event in the mitigation drive compared with the baseline drive.

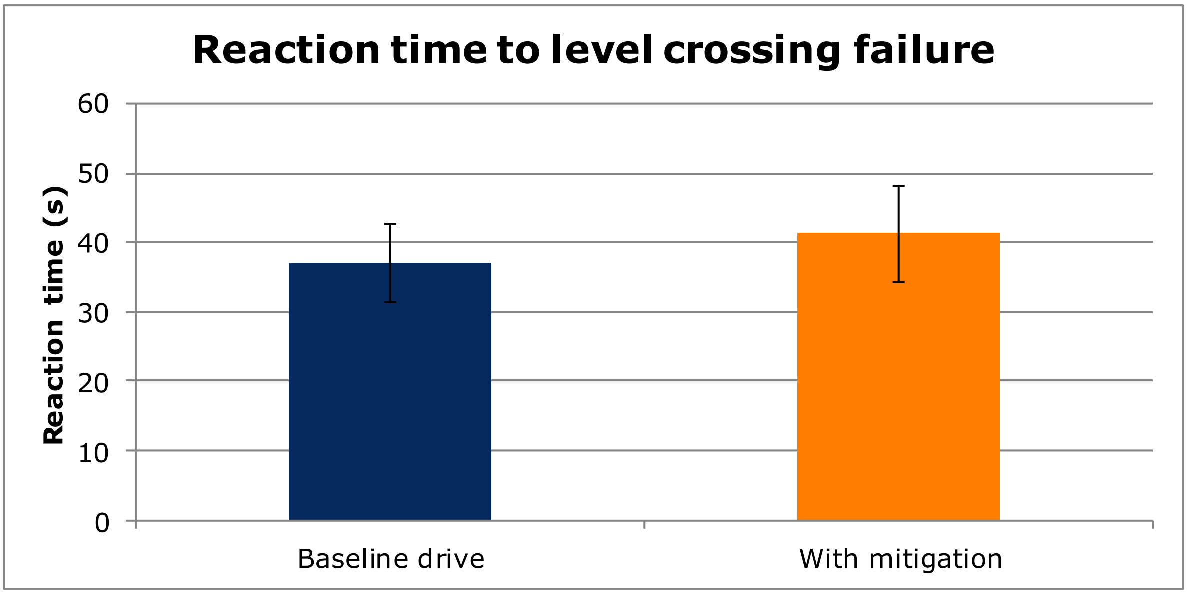


Figure 8: Reaction time to level crossing failure

Figure 8 shows that the reaction time to the level crossing failure. A paired samples t-test showed that the differences were not statistically significant (t(7) = .468, p=.653). The raised barriers at the level crossing were noticed by all nine participants.

6.1.8 DSD pedal activations

Hypothesis H08 was that participants would perform fewer driver safety device (DSD) activations in the mitigation drive compared with the baseline drive.

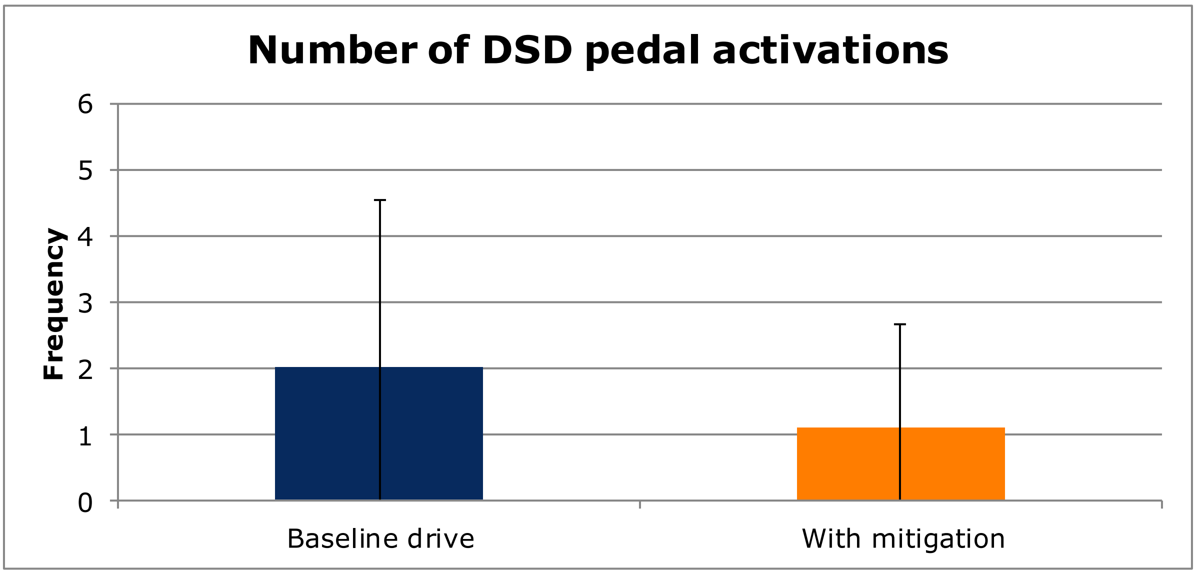


Figure 9: Frequency of DSD activations

Figure 9 displays the average number of times the driver safety device was triggered, requiring drivers to release and re-engage DSD pedal. A Wilcoxon Signed Ranks Test showed that the differences were not statistically significant (Z= -1.382, p=.167).

6.2 Subjective measures

6.2.1 *NASA TLX*

Hypothesis H09 was that after the mitigation drive, participants would report higher mental workload on the NASA-TLX compared with the baseline drive.

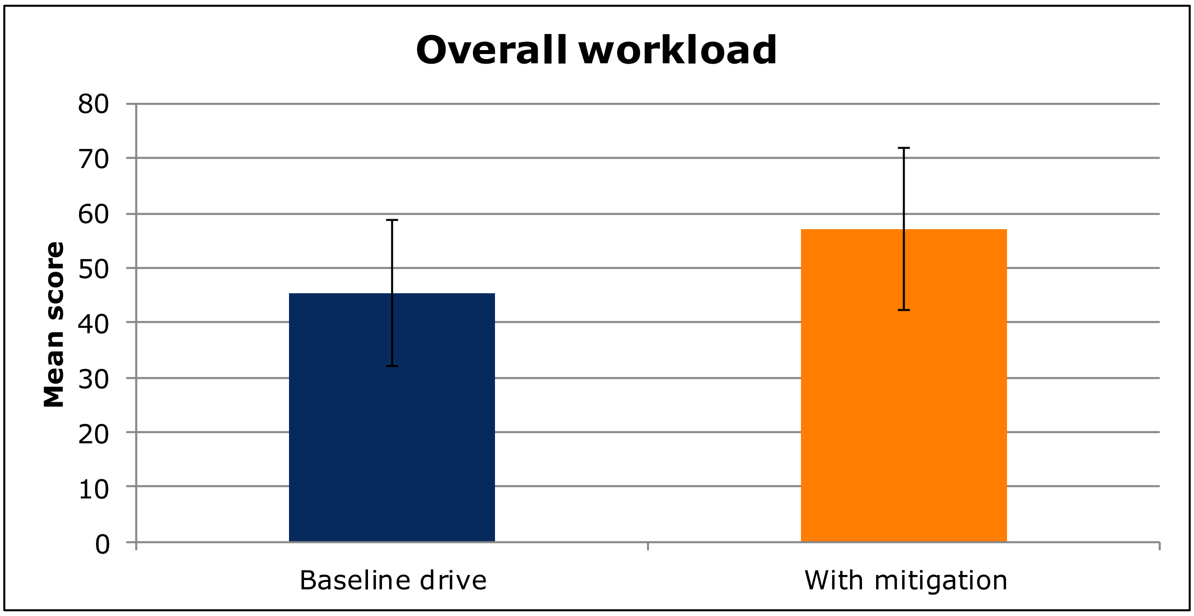


Figure 10: Overall perceived workload

Figure 10 displays the mean overall workload ratings for the nine participants in both conditions. The perceived workload after the mitigation drive was considerably higher than after the baseline drive. Paired samples t-test confirmed that the differences were statistically significant ($t(8) = -2.571$, $p=0.033$).

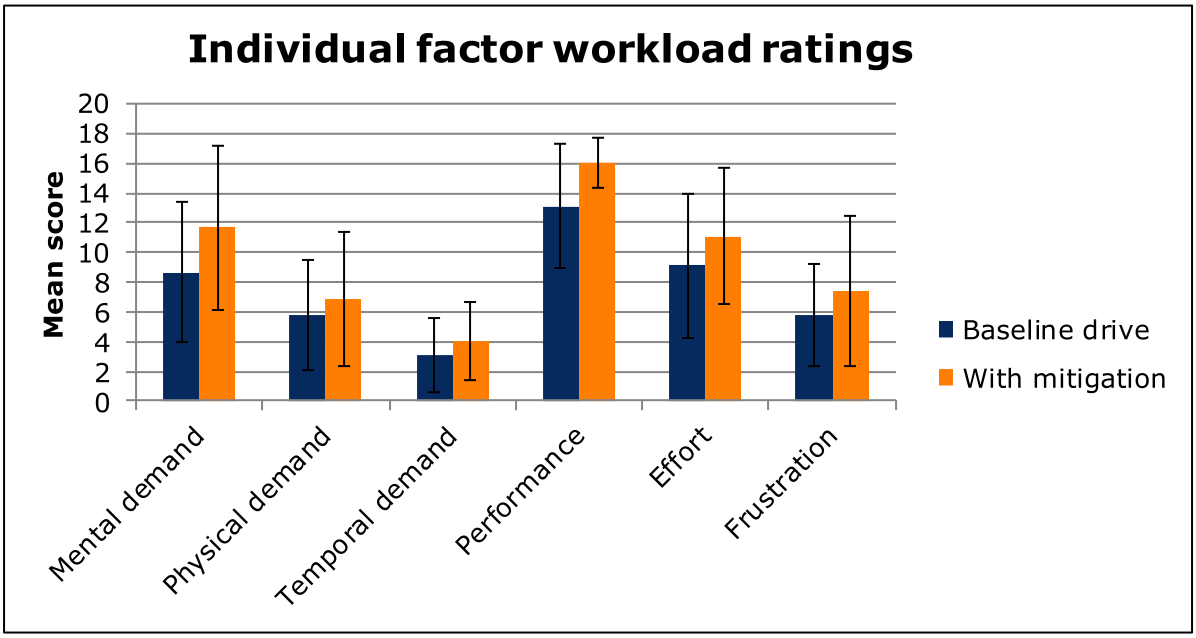


Figure 11: Individual NASA TLX workload ratings

Figure 11 shows a breakdown of the individual NASA TLX ratings. Paired samples t-tests showed that the differences in the performance (t(8) = -3.043, p=.016) and effort (t(8) = -1.973, p=.084) factors were statistically significant, with drivers reporting higher perceived performance and effort when the mitigation was applied.

6.2.2 Karolinska Sleepiness Scale (KSS)

Hypothesis H10 was that participant's perceived levels of fatigue would be lower in the mitigation drive compared with the baseline drive. The effect of time on task was also evaluated by taking measurements before, during and after each drive.

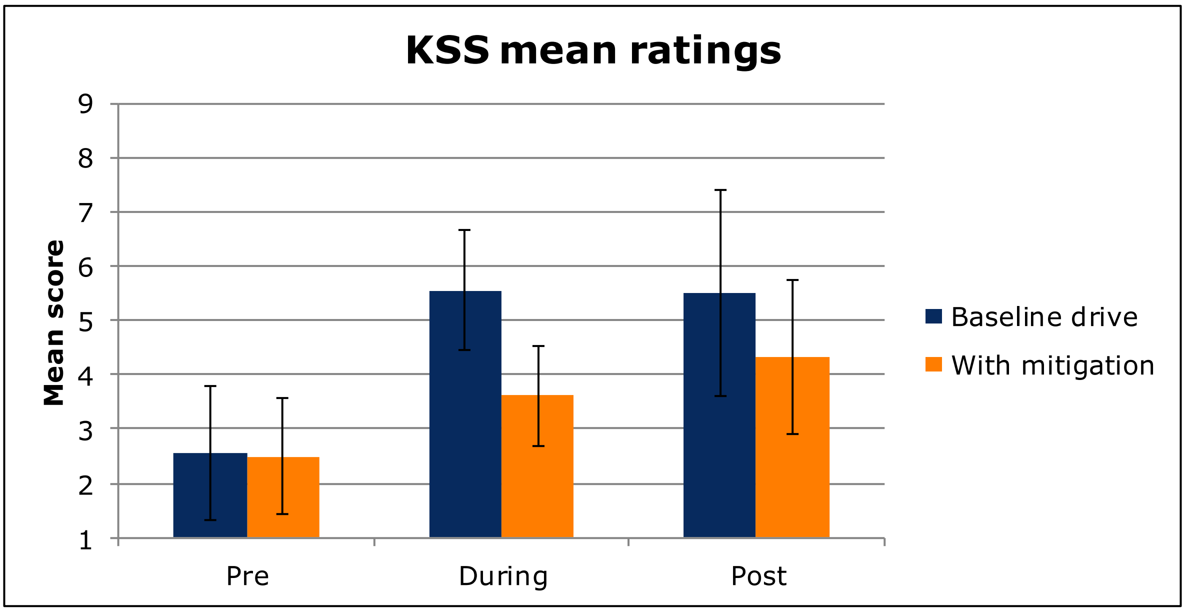


Figure 12: KSS mean ratings

Figure 12 shows that the mean sleepiness ratings increased throughout both drives, but were lower in the mitigation condition both during and after the drive. This shows that there was a delayed onset of sleepiness in the mitigation condition. A two-factor repeated measures ANOVA was carried out and showed that there was a main effect of time on task on reported sleepiness (F(1, 2) = 28.513, p=.000), and that the effect of the mitigation was approaching statistical significance (F(1,1) = 3.158, p=.113). There were no statistically significant differences between the two drives in terms of the KSS scores taken before the drive. This suggests that participants were in a similar state before each drive, and that sleepiness was not a confounding variable.

6.2.3 Stress Arousal Checklist (SACL)

Hypothesis H11 was that after the mitigation drive, participants would report higher arousal levels on the SACL compared with the baseline drive.

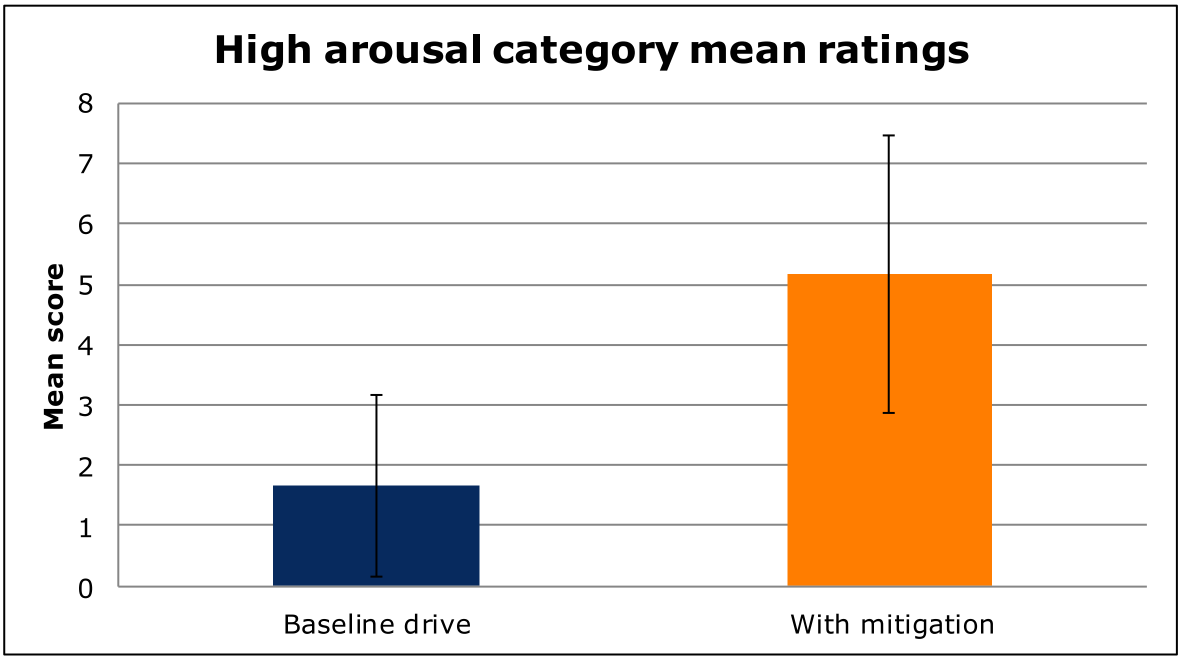


Figure 13: Stress arousal check list- high arousal mean ratings

Figure 13 presents the mean scores from the high arousal category of the stress arousal checklist administered to the nine participants. Data was not normally distributed so a Wilcoxon Signed Ranks test was conducted. From the analysis it was found that there was a difference in high arousal scores between both conditions ($Z = -2.060$, $p = .039$).

Table 4: The frequency of 'yes' and 'no' answers to each adjective on the stress arousal checklist with and without the mitigation

Category	Adjective	Baseline drive		With mitigation	
		No	Yes	No	Yes
High stress levels	Worried	9	0	9	0
	Fearful	8	1	8	1
	Dejected	8	1	9	0
	Up-tight	9	0	8	1
	Uneasy	8	1	7	1
	Jittery	9	0	8	1
	Bothered	9	0	9	0
	Apprehensive	6	3	6	3
	Tense	9	0	6	2
	Nervous	7	2	7	1
	Distressed	9	0	8	1
Low stress levels	Peaceful	1	8	1	8
	Calm	0	9	0	9
	Contented	0	8	1	7
	Cheerful	3	5	1	7
	Pleasant	1	7	0	9
	Comfortable	0	9	0	9
	Restful	1	8	1	7
	Relaxed	0	9	0	9
High arousal levels	Active	7	2	1	8
	Activated	5	4	1	8
	Stimulated	8	1	2	7
	Vigorous	7	0	6	1
	Alert	1	8	0	9
	Lively	7	2	4	4
	Aroused	7	1	5	2
	Energetic	6	1	2	4
Low arousal levels	Drowsy	1	8	6	3
	Idle	4	5	7	2
	Tired	2	5	5	4
	Sleepy	1	7	8	1
	Sluggish	4	5	9	0
	Somnolent	3	4	3	2

The frequency of 'yes' and 'no' answers to each factor are shown in **Table 4** for the baseline and mitigation drives. Key changes are shown in green. The results show that after the simulator drive with the mitigation more respondents said that they were tense, cheerful, activated, lively and energetic. Fewer respondents said that they were drowsy, idle, tired and sluggish.

6.2.4 Interview response

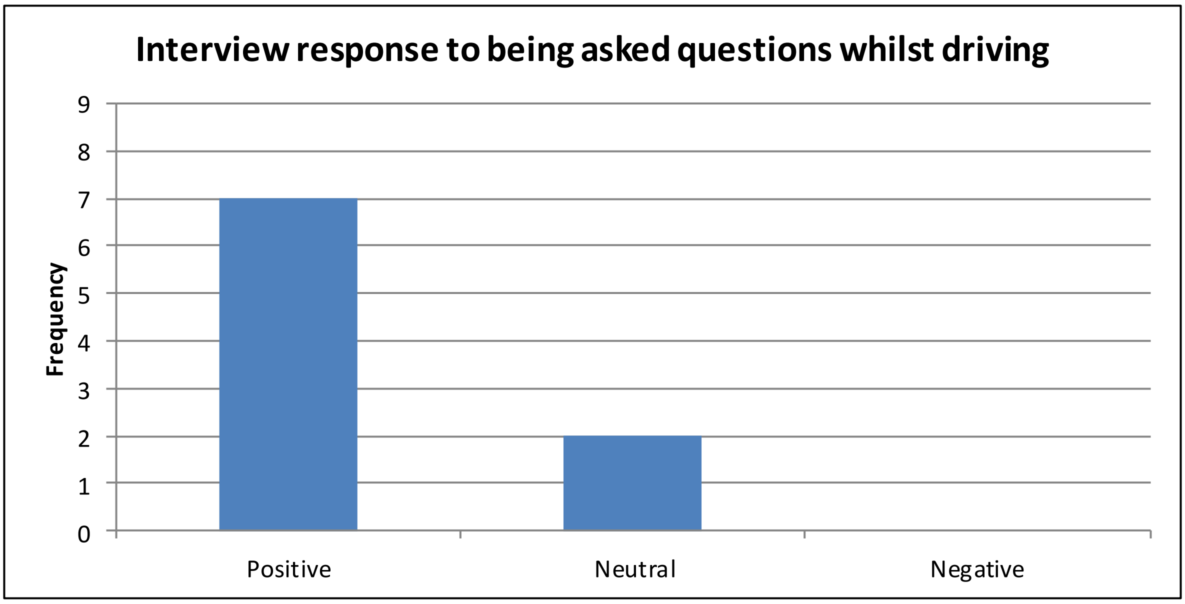


Figure 14: Interview response to being asked questions

Figure 14 compares the response of the nine participants when asked in the interview: 'overall what impact do you think the questions had on your driving-positive/negative/neutral?' The graph shows a clear positive response from the participants.

The following themes emerged from the interview:

- There was a generally positive response to the mitigation questions.
- The mitigation questions helped keep participants more focused on the task.
- Participants felt that the mitigation questions helped prevent feelings of boredom and increased their alertness levels.
- There was mixed response as to whether the mitigation questions affected driving performance.
- The frequency and difficulty of the mitigation questions were reported to be appropriate for the task, as was the terminology used in the question set.
- Participants were clear that they would not want the mitigation questions to be used in a real life driving scenario.
- Participants felt that other drivers were also unlikely to find the mitigation questions acceptable in a real-world scenario.

6.3 Physiological measures

6.3.1 Heart rate

On examination of the data collected it appeared that there were a number of abnormalities, with clearly erroneous readings as high as 2136bpm. Several steps were taken in order to cleanse the data. Error corrections were applied using the Polar software to remove the abnormal heart rate readings. However, even after these steps were taken there remained significant concerns about data quality. For this reason, analysis of the data was not taken further.

It is not clear why such problems were experienced with the heart rate data. Normally users of such equipment would be undertaking physical activity and producing sweat, thus keeping the electrodes moist and allowing them to function correctly. This was not the case in this study and may have been a factor affecting data quality.

7 Discussion

This study was conducted to identify the types of task and situational factors that might lead to cognitive underload and reduced performance during driving, to investigate potential mitigations and trial one such measure in a train driving simulator.

The literature review highlighted that workload is related to arousal. If a person's workload is low, this can lead to narrowed attention or alertness, reduced motivation and poorer performance. It also revealed that low workload tasks have certain characteristics. These characteristics create tasks that are monotonous, repetitive, potentially long in duration, and have an objectively low level of work or difficulty associated with the task.

This study carried out a desk-based assessment of train driving tasks against the range of task characteristics associated with low workload, which were highlighted in the literature review. The following train driving tasks were highlighted, through this process, as having characteristics associated with low demand:

- Evaluating the speed that the train should be driven, based on speed restrictions, signal aspects, route knowledge, etc.
- Driving according to the speed strategy.
- Reading signals.
- Ensuring the train is in safe mode, when the train has stopped at a station.
- Checking platform end signal for the Close Doors Indicator during platform staff assisted dispatch.
- Driver-guard station duties.

The study also identified a number of situational factors that could affect workload, including:

- Design and layout of information, such as schedule cards, operating notices. The more difficult they are to read the higher the demand of the task.
- The diagram, and in particular diagrams that are repetitive, or predictable, those that do not require the driver to intervene very much to maintain safe operation, and long diagrams.
- Familiarity with the diagram, location or rolling stock.
- The driver's stress levels and emotional response to the situation (although in most cases it is likely that the driver will be calm and not under stress).
- The environment, particularly those that are predictable, repetitive and lack distractions.
- Rolling stock/cab design, including comfort and whether the driver is able to stand up.
- Type of operation (i.e. whether driver workload is reduced by allocation of tasks to others such as guards).
- Absence of communication.

Based on this information, the simulator study used a scenario where the driver was required to drive for a long period of time on green signals, without stopping at any stations.

The aim of the simulator study was to evaluate whether application of a mitigation against cognitive underload would have a measureable effect. A focus group was therefore convened to identify a number of potential techniques that could be used to mitigate against underload. A number of different options were discussed, some of which could be strong candidates for further development in other studies. Due to the small scale of the current study, the aim was to choose a measure that could be easily implemented within the context of a simulator study, and not necessarily the real world. The mitigation selected was a series of verbal questions about the status of the train and the environment. These questions were designed to be quick and easy to answer. They were delivered one at a time by the experimenter at fixed intervals of one mile (which broadly equates to one every two minutes).

Nine participants took part in the simulator study on Southern's train driving simulator in Selhurst. Participants were from a range of train operating companies, and drove the same simulated route on two occasions. On one occasion baseline measures of driving performance, subjective response and heart rate were taken. On a separate occasion the same measurements were taken, but the mitigation described above was applied in order to increase workload.

Contrary to the hypotheses, the results of the experiment did not show that the participants' driving performance varied significantly between the baseline and mitigation conditions. It is not clear why this was the case, though some suggestions are offered towards the end of this section. Nevertheless, the subjective measures of workload, fatigue and arousal suggest that the mitigation had a positive impact on how participants felt.

Self-reported workload (based on the NASA TLX) increased when the mitigation was introduced, when compared with the baseline. The overall mean score increased from 45.6 after the baseline drive to 57.1 after the mitigation drive. It is interesting that two of the NASA TLX's component factors, perceived performance and effort, were also higher after the mitigation drive, when compared with the baseline. These findings are consistent with previous research by Chan and Atchley (2011) who indicated that a verbal task can increase effort.

Participants' sleepiness levels were measured using the Karolinska Sleepiness Scale (KSS). Encouragingly, mean sleepiness levels before both drives were very similar, suggesting that sleepiness before the drive was not a confounding factor in this study. Sleepiness increased during both the baseline and mitigation conditions. The rating for the non-mitigation drive increased rapidly after the pre-drive rating and then levelled at a mean rating of 5.5 (beginning to show signs of sleepiness). The rating for the mitigation drive also increased but at a slower rate, reaching a mean rating of 4.3. Thus, the onset of sleepiness was delayed in the mitigation drive. Previous research by Dunn (2011) also found that monotonous tasks can lead to increases in feelings of sleepiness and tiredness.

Larue, Rakotonirain and Pettitt (2011) found that low workload reduces the individual's arousal levels. Analysis of the stress arousal checklist completed by participants showed that there was an increase in scores for high arousal factors following the mitigation drive (5.2) when compared with the baseline (1.7). Following the mitigation drive, participants felt more tense, cheerful, activated, lively and energetic, and less drowsy, idle, tired and sluggish. The aim of the mitigation was to increase the level of cognitive load on the driver. An effect of this increase (as identified by previous research by Hancock and Warm, 1989) is an improved ability to maintain an optimum level of arousal. The increase in arousal levels during the mitigation drive, as shown by the stress arousal checklist, supports this theory and suggests the mitigation had the desired effect in this respect.

Disappointingly there were issues with heart rate data quality and so the data was not analysed. A large amount of noise was present in the results which may have been caused by the lack of moisture on the electrodes. The Polar heart rate monitor is designed to be worn when exercising and the contact between the monitor and the chest relies on moisture to improve the conductivity of electrical signals produced by the heart. Sweating improves the accuracy of readings. Although the elastic strap was moistened at the start of the experiment, it is likely that it dried out due to the nature of the low workload experiment. Heart rate seems to be highly affected by a range of different factors and although it is easy to administer, it may not produce the most accurate results.

Given the effect of the mitigation on how participants felt, it is interesting that the study was unable to detect any differences in objective performance measures such as speed control and response to critical events. There are a number of reasons why this may have been the case. Firstly, the literature (e.g. Dunn, 2011; Hockey, 1997) highlights motivation and effort as being a potential buffer against the detrimental effects of low workload. In other words, even if task demands are very low, motivated individuals may succeed in remaining alert for some time. Subjective feelings such as sleepiness, tiredness or boredom precede any performance effects. Thus, perhaps the participants were successful in warding off cognitive underload in both of the drives, and as a result, maintained an acceptable level of performance.

A second potential reason why the performance measures showed no difference is that the scenario may not have been easy and familiar enough. One of the likely contributory factors to low workload is good route knowledge (Naghiyev, 2014). If the driver is very familiar with the route he is driving then the issues identified previously such as lowered arousal levels and decreased workload levels may become increasingly prevalent. The majority of the drivers who took part in the study had never driven the simulator before and had never experienced the route used in the experiment. The unfamiliar route may have increased workload, and prevented a state of cognitive underload from being reached. Indeed, in the interviews one participant said, 'I feel that the unfamiliar trip meant I was slightly more alert than I normally would be'. The class 377 simulator itself was also an unfamiliar type of rolling stock for some of the drivers. Although participants were familiar with the general layout of key controls of the 377 simulator, they did not have the traction knowledge for that type of train. These factors may have further raised participants' workload. Some participants' lack of initial familiarity with the location of the hill start button may have had a similar effect, even though participants were briefed on the location of this button. The hill start button was used by participants in this study that they had detected an irregularity, such as the lack of an AWS bell.

The final reason why no differences were found could be that the effect of underload was present but just not detected due to compromises made during the study design. A training simulator was used for this study rather than a research simulator. The scope for altering the training simulator route and scenario was very limited, especially when it came to creating critical events. This may have affected the quality of the reaction frequency and reaction time data. Individual differences may have played a major role in the reaction frequency measures associated with the AWS irregularities. The data suggests that participants were either sensitive to these events or not. This explains the large number of AWS irregularities missed in both conditions as shown in figure 7. In addition, few significant bends were present in the drive meaning events such as the level crossing irregularity were difficult to hide, instead appearing on a long straight stretch of railway. This may have contributed to the highly variable and long response times as reported in the results. Responding to these events may have been a test of how well the participants could interpret a pixelated display of an object in the distance and may have related more to their visual abilities. The simulator was also not capable of producing reaction time measurements in as much granularity as had initially been hoped.

It is clear to see from the interview data that the general response to the mitigation was positive. Participants generally thought that the mitigation technique helped keep them alert and focused on the task of driving. The majority of the nine participants identified that the effect (positive or negative) of this particular mitigation would heavily depend on the individual. "Other drivers have differing styles, others might not want it [the mitigation] as it could affect their performance." This is an interesting finding and might relate to many mitigation techniques against low demand. Another theme was the mention that older drivers would be against this particular mitigation. When asked about the effect the questions had on specific emotions, the general response was that they helped prevent boredom and increase alertness levels.

8 Conclusions

In conclusion, the subjective measures from the study seem to suggest that applying a mitigation which increases workload during an undemanding train driving scenario has a positive impact on self-reported measures of low workload, low arousal and fatigue. Further study needs to be carried in order to understand if this effect would persist over time.

There are a number of potential reasons why the study did not find an effect on task performance. It is possible that motivated participants may have successfully kept underload at bay. On the other hand, the outcome may have been a result of compromises in study design. Future work should, if possible, consider easier, familiar scenarios, optimise the design of critical events and take more granular measurements. Conducting longer-term studies may also be important in overcoming the potential effect of novelty on participant motivation and therefore performance.

The physiological measure of heart rate did not yield data of sufficient quality for analysis. As explained in chapter 7, the large amount of noise in the data was possibly due to the heart rate monitor equipment being designed primarily for use during exercise. In future research it is important to take into consideration the primary purpose of the equipment and conduct a more rigorous pilot study if equipment is being used for a different purpose.

As was expected, participants thought that the mitigation technique had positive effects, but were quick to point out that it is not suitable for adoption in the real world. If the mitigation was to be taken further, ways to make the technique adaptive to the journey and driver would need to be considered, in order to ensure that it does not overload drivers in high workload situations. Further research needs to be conducted to develop and test measures that are effective and safe to use in the real world.

9 References

Barton, T. (2003). On the safe side. Overdrive.

Benedetto, S., Pedrotti, M., Minin, L., Baccino, T., Re, A., Montanari, R. (2011). Driver workload and eye blink duration, Transportation Research Part F-traffic Psychology And Behaviour. 14:199-208.

Brodsky, W., Kizner, M. (2012). Exploring an alternative in-car music background designed for driver safety, Transportation Research Part F-traffic Psychology And Behaviour. 15:162-173.

Brown, I.D. (1994). Driver fatigue. Human Factors, 36, 298-314.

Buksh, A., Sharples, S., Wilson, J.R., Coplestone, A. and Morrisroe, G. (2013). A comparative cognitive task analysis of the different forms of driving in the UK rail system. Rail Human Factors: Supporting reliability, safety and cost reduction. Boca Raton, FL: CRC Press

Casali, J.G., Wierwille, W.W. (1984). On the measurement of pilot perceptual workload: A comparison of assessment techniques addressing sensitivity and intrusion issues. Ergonomics. 27:1030-1050.

Chan, M., Atchley, P. (2011). Potential Benefits and Costs of Concurrent Task Engagement to Maintain Vigilance: A Driving Simulator Investigation, Human Factors. 53:3-12.

Curcio, G., Casagrande, M., Bertini M. (2001). Sleepiness: evaluating and quantifying methods, International Journal of psychophysiology. 41:251-263.

Damrad-Frye, R. & Laird, J.D. (1989). The experience of boredom: The role of the self-perception of attention, Journal of Personality & Social Psychology. 57, 315-320.

Deaton, J., Parasuraman, R. (1997). Effects of task demands and age on vigilance and subjective workload. Proceedings of the Human Factors and Ergonomics Society Annual meeting. 32: 1458-1462.

Desmond, P.A. & Hancock, P.A. (2001). Active and passive fatigue states. In P.A. Hancock & P.A. Desmond (Eds.), Stress, Workload, and Fatigue (pp. 455-465). New Jersey: Lawrence Erlbaum.

De Valck, E., Cluydts, R. (2003). Sleepiness as a state–trait phenomenon, comprising both a sleep drive and a wake drive, Medical Hypothesis. 60:509-512.

Dinges, D. F., Mallis, M., Maislin, G., Powell, J. W. (1998). Final report: Evaluation of techniques for ocular measurement as an index of fatigue and as the basis for alertness management, National Highway Traffic Safety Administration.

Dunn, N. J. (2011). Monotony: The effect of task demand on subjective experience and performance, School of Risk and Safety Sciences, University of Wales.

Dunn, N., Williamson, A. (2012). Driving monotonous routes in a train simulator: the effect of task demand on driving performance and subjective experience, Ergonomics. 55:997-1008.

Driver, J. (2001). A selective review of selective attention research from the past century, British Journal of Psychology. Vol.92(1), pp.53-78

Easterbrook, J.A. (1959). The effect of emotion on cue utilization and the organization of behaviour, Psychological Review. 66, 187-201.

Edkins, G.D. & Pollock, C.M. (1997). The influence of sustained attention on railway accidents. Accident Analysis and Prevention. 29, 533-539.

Fahlman, S.A., Mercer-Lynn, K.B., Flora, D.B., Eastwood, J.D. (2013). Development and validation of the multidimensional state boredom scale, PubMed. 20:68-85.

Fairclough, S.H. (2001). Mental effort regulation and the functional impairment of the driver, Stress, Workload, and Fatigue (pp. 479-502). New Jersey: Lawrence Erlbaum.

Gershon, P., Ronen, A., Oron-Gilad, T., & Shinar, D. (2009). The effects of an interactive cognitive task (ICT) in suppressing fatigue symptoms in driving. Transportation Research Part F, 12, 21-28.

Grandjean, E. (1979). Fatigue in industry. British Journal of Internal Medicine, 36, 175-186.

Hancock, P.A. & Warm, J.S. (1989). A dynamic model of stress and sustained attention. Human Factors, 31, 519-537.

Hart, S.G. & Staveland, L.E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), Human Mental Workload (pp. 139-183). North Holland: Oxford.

HEL (2004). Early human factors analysis for UK-ERTMS driving (final version). Accessed online 12th November 2013: www.sparkrail.org

Hockey, G.R.J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. Biological Psychology. 45, 73-93.

Hitchcock, E.M. & Matthews, G. (2005). Multidimensional assessment of fatigue: A review and recommendations. In Proceedings of the International Conference on Fatigue Management in Transportation Operations. Seattle, USA.

Jellentrup, N., Metz, B., Rothe, S. (2011). Can talking on the phone keep the driver awake? Results of a field-study using telephoning as a countermeasure against fatigue while driving. Proceedings of the 2nd International Conference on Driver Distraction and Inattention.

Kahneman, D. (1973). Attention and Effort. New Jersey: Prentice-Hall.
Kobayashi, S., Hara, T., Goi, Y. (1996). Predicting changes in driver's drowsiness level during prolonged driving: extracting long blinks from EOG data. JSAE Review. 17, 435–458.

Lal, S.K.L. & Craig, A. (2001). A critical review of the psychophysiology of driver fatigue. Biological Psychology, 55, 173-194.

Larue, S.G., Rakotonirainy, A., Pettitt, A.N. (2011). Driving performance impairments due to hypovigilance on monotonous roads, Accident Analysis & Prevention. 43: 2037-2046.

Lavie, N. (1995). Perceptual Load as a Necessary Condition for Selective Attention, American Psychological Association. 21: 451-468

Lavie, N., Lin, Z., Zokaei, N., & Thoma, V. (2009). The Role of Perceptual Load in Object Recognition. Journal of Experimental Psychology. Human Perception and Performance, 35(5), 1346–1358.

MacDonald, W. (2003). The Impact of Job Demands and Workload on Stress and Fatigue, Australian Psychologist. 38:102–117

Mackie, R.R., Wylie, C.D. and Smith M.J. (1994). Countering loss of vigilance in sonar watchstanding using signal injection and performance feedback. Ergonomics, 37(7), p1157-1184.

Mallis, M. M. (1999). Evaluation of techniques for drowsiness detection: Experiment on performance-based validation of fatigue-tracking technologies. Unpublished Doctoral Thesis, Drexel University, Philadelphia, PA.

Matthews, G. & Desmond, P.A. (2002). Task-induced fatigue states and simulated driving performance. The Quarterly Journal of Experimental Psychology, 55A, 659-686.

May, J.F. & Baldwin, C.L. (2009). Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. Transportation Research Part F, 12, 218- 224.

McBain, W.N. (1970). Arousal, monotony, and accidents in line driving. Journal of Applied Psychology, 54, 509-519.

Merat, N., Carsten, O., Lai, F.C., Barnard, Y., Jamson, A.H. (2012). Control task substitution in semiautomated driving: does it matter what aspects are automated? Human Factors. 54: 747-761.

Merat, N., Jamson, A.H., (2013). The effect of three low-cost engineering treatments on driver fatigue: A driving simulator study, Accident Analysis and Prevention. 50: 8.

Naghiyev, A., Sharples, S., Carey, M. and B. Ryan. (2014). ERTMS train driving – incab vs outside: an explorative eye-tracking field study. In S. Sharples and S.T. Shorrock (Eds.) Contemporary Ergonomics and Human Factors. London: Taylor and Francis

Nachreiner, F (1995). Standards for ergonomics principles relating to the design of work systems and to mental workload. Applied Ergonomics, 26, 259-263.

Neubauer, C., Matthews, G., Saxby, D. (2012). The effects of cell phone use and automation on driver performance and subjective state in simulated driving, Proceedings of the Human Factors and Ergonomics Society. 1987-1991.

Oken, B.S., Salinsky, M.C., & Elsas, S.M. (2006). Vigilance, alertness, or sustained attention: Physiological basis and measurement. Clinical Neurophysiology, 117, 1885-1901.

Parasuraman, R., Molloy, R., Indramani, L.S. (1993). Performance Consequences of Automation-Induced 'Complacency. The International Journal of Aviation Psychology. 3: 1-23.

Porr, C., Olson, K., Hegadoren, K. (2010). Tiredness, Fatigue, and Exhaustion in the Context of a Major Depressive Disorder, Qualitative Health Research. 20:1315-1326.

Razmjou, S. (1996). Mental workload in heat: Toward a framework for the analyses of stress states. Aviation, Space, & Environmental Medicine, 67, 530-538.

Rail Safety and Standards Board (2005). Mental workload assessment for train drivers (T147). Accessed online 10th October 2013: www.sparkrail.org

Sarter, M., Givens, B., Bruno, J.P. (2001). The cognitive neuroscience of sustained attention: where top-down meets bottom-up, Brain Res Rev. 35(2):146-60.

Saxby, D.J., Matthews, G., Hitchcock, E.M., & Warm, J.S. (2007). Development of active and passive fatigue manipulations using a driving simulator. In Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting, Santa Monica, USA.

Shapiro, C.M., Auch, C., Reimer, M., Kayumov, L., Heslegrave, R., Huterer, N., Driver, H., & Devins, G.M. (2006). A new approach to the construct of alertness. Journal of Psychosomatic Research, 60, 595-603.

Straussberger, S. (2006). Monotony in Air Traffic Control: Contributing Factors and Mitigation Strategies, Eurocontrol Experimental Centre.

Straussberger, S. & Schaeffer, D. (2007). Monotony in air traffic control. Air Traffic Control Quarterly, 15, 183-207.

Thayer, R.E. (1978). Towards a psychological theory of multidimensional activation (arousal). Motivation and Emotion, 2, 1-34.

Thornburg, K.M., Peterse, H.P.M., Liu, A.M. (2012). Operator performance in long duration control operations: Switching from low to high task load, Proceedings of the Human Factors and Ergonomics Society. 2002-2005.

Thiffault, P. & Bergeron, J. (2003). Monotony of road environment and driver fatigue: A simulator study. Accident Analysis and Prevention, 35, 381-391.

Unal, A.B., Platteel, S., Steg, L., Epstude, K. (2013). Blocking-out auditory distracters while driving: A cognitive strategy to reduce task-demands on the road, Accident Analysis and Prevention. 50: 934-939.

Verwey, W.B., Zaidel, D.M. (1999). Preventing drowsiness accidents by an alertness maintenance device, Accident Analysis And Prevention. 31: 199-211.

Wilde, G.J.S. & Stinson, J.F. (1983). The monitoring of vigilance in locomotive engineers. Accident Analysis and Prevention, 15, 87-93.

Williamson, A., Lombardi, D.A., Folkard, S., Stutts, J., Courtney, T.K., & Connor, J.L. (2011). The link between fatigue and safety. Accident Analysis and Prevention, 43, 498-515.

Young, M.S. & Stanton, N.A. (2002a). Malleable attentional resources theory: A new explanation for the effects of mental underload on performance. Human Factors, 44, 365-375.

Young, M.S. & Stanton, N.A. (2002b). Attention and automation: New perspectives on mental underload and performance. Theoretical Issues in Ergonomics Science, 3, 178-194.

Young, M.S. & Stanton, N.A. (2006). In loco intellegentia: Human factors for the future European train driver. International Journal of Industrial and Systems Engineering, 1, 485-501.

Zheng, B., Jiang, X.T., Tien, G., Meneghetti, A., Panton, O.N.M., Atkins, M.S. (2012). Workload assessment of surgeons: correlation between NASA TLX and blinks, Surgical Endoscopy And Other Interventional Techniques. 26: 2746-2750.

Appendix A Mitigation measures considered

Measure	Origin	Evaluation outcome	Comments
Interactive secondary task – quiz/game – presented at low workload times. Driver chooses to engage or not. Verbal presentation and response.	Lit review	Chosen as mitigation to trial	Interactive secondary tasks were considered to be promising candidates. Possible difficulties around keeping the question set engaging in the long term. For the current trial, questions probing situational awareness were used.
Interactive secondary task – game – trackside shoot 'em up with competitive element	Workshop	Combined with adding extra critical stimuli	-
Interactive secondary task – game – tetris type game with visual presentation and manual response and competitive element	Workshop	Discounted during workshop	Interactive secondary tasks were considered to be promising candidates. Concerns around visual distraction if attention is directed away from where it is required for driving.
Monitoring of external environment (e.g. photography of infrastructure problems, foliage, graffiti, etc.)	Workshop	Discounted during workshop	Driver should already monitor infrastructure and report any obvious issues. Possible difficulties in motivation to carry on looking for issues if the condition of infrastructure is not improved after reporting.
Projecting false critical stimuli onto the visual scene (e.g. red or amber signals requiring an acknowledgement)	Workshop	Discounted during pilot study	Inspired by Artificial Signal Injection (ASI), which involves the random presentation of false target signals during a vigilance task. It increases the frequency with which operators encounter target signals and changes their expectations regarding the frequency/timing of target appearance. This was found to significantly improve detection rates for sonar signals when combined with feedback (Mackie et al., 1994). More recently the concept of ASI has been used in Treat Image Projection (TIP) systems for baggage x-ray scanners, which randomly project images of fictional threat items such as weapons, among images of real passenger possessions. Promising as a measure but would require significant resource to implement in the simulator.
Adding a physical element to the task (e.g. pedalling)	Workshop	Discounted during workshop	Increases physical demand. Unlikely to be acceptable on longer journeys.

Measure	Origin	Evaluation outcome	Comments
Listening to the radio	Lit review	Combined with the idea about playing music	-
Playing music (standardised, neutral music that would not mask alarms and alerts)	Workshop	Discounted following pilot study	This was thought to be a potentially promising measure which would be easy to implement, and was a close contender. However, playing music in the cab has been discussed within the rail industry before and is not considered to be an acceptable measure.
Timing food and drink intake to optimise sugar levels. Potential reward element.	Workshop	Discounted during evaluation	
Driver briefing, workload monitoring (rating scale) task and commentary driving when ratings indicate underload.	Workshop	Discounted during evaluation	Potentially promising and easy to implement, but in the context of a short simulation study this measure would amount to telling people to try harder during one of the drives. It would be difficult to demonstrate the measure works without a longer term study.
Occluding the visual scene (e.g. with fog in the simulator) periodically. This would be done rarely and at non-critical times.	Workshop	Discounted during evaluation	Although this is likely to be effective it is not clear how easy it would be to implement in the simulator, and it might be difficult to implement/ unacceptable in the real world.
Pre-journey briefing and team discussion session in which the route and risks are discussed.	Workshop	Discounted during workshop	Potentially useful as a motivator but would require a culture change. Effectiveness might not be as high as some of the other measures.
Driver briefing and workload monitoring (rating scale) task.	Workshop	Combined with the commentary driving measure	-
Introducing smells into the cab (specific smells associated with particular signal aspects, and other smells at low workload times).	Workshop	Discounted during workshop	Difficult to achieve such that smells enter the cab swiftly and then leave swiftly. Previous research suggests that rousing effects may not last very long.
Increasing task difficulty (providing numerical DAS information as a sum of two numbers rather than a number, so that the driver has to calculate the advised speed).	Workshop	Discounted during workshop	Potential for miscalculation.
Feedback on the status of the train (eg performance, timing, smoothness) – with competitive potential.	Workshop	Discounted during workshop	Could be perceived as promoting performance measures over safety.

Measure	Origin	Evaluation outcome	Comments
Making it harder to maintain set speeds by programming the train speed to drift.	Workshop	Discounted during workshop	Hard to achieve in a way that takes into account driver workload.
Use of Variable Message Signs at known problem locations along routes. Driver would be required to indicate what the message was (like a quiz).	Lit review	Discounted during workshop	Considered to be quite similar to the interactive secondary tasks.
Conversing on a mobile phone.	Lit review	Not evaluated	Company policies are unlikely to be changed to allow use of mobile phones.
Modifying the working conditions to make them less ‘sterile’.	Lit review	Not evaluated	Hard to target low workload situations only.

Appendix B Verbal secondary task mitigation questions

1. What is the current main reservoir pipe reading?
2. Please give one situation in which you would set the driver reminder appliance.
3. What time is currently displayed on your Mitrac screen?
4. What aspect was the last signal you passed displaying?
5. If you were calling a signaller to report an emergency, how would you start the conversation, in order to convey the urgency of the situation?
6. What notch is your power brake controller currently in?
7. Have you driven over any level crossings on this drive?
8. If you are standing lineside and the driver of an approaching train sounds the horn, what must you do?
9. Do the signals in this area have three aspects or four aspects?
10. What is the current speed of your train?
11. Using the phonetic alphabet, please tell me your train running number.
12. When you are at a station and you receive a bell code of 1-2 from the conductor, what does this mean?
13. What was the name of the last station you passed?
14. Name one piece of equipment that a train driver must have with them when they are on duty.
15. What is the current main reservoir pipe reading?
16. How many tunnels have you driven through on this drive?
17. Please give one possible reason why your train might be brought to a stand by a brake application which you did not make?
18. What time is currently displayed on your Mitrac screen?
19. What aspect was the last signal you passed displaying?
20. Between what times of day should a train driver sound the horn when passing a whistle board?
21. What notch is your power brake control currently in?
22. How many level crossings have you driven over on this drive?
23. If a train driver is asked to proceed at caution, what does this mean?
24. Do the signals in this area have three aspects or four aspects?
25. What is the current speed of your train?
26. If the passcom is activated on a train you are driving, name one type of location where you would normally want to avoid stopping the train.
27. What is the current brake pipe reading?

28. If you are at a station being dispatched by platform staff and you see an indication with the letters RA, that is Romeo Alpha, what does that mean?
29. When preparing a train for service, how many detonators should be available on board?
30. What was the name of the last station you passed?
31. Name one piece of equipment you can operate to reset the driver vigilance device?
32. What time is currently displayed on your Mitrac screen?
33. What aspect was the last signal you passed displaying?
34. What would you do if you saw a trackworker standing lineside in a place of safety?
35. What notch is your power brake control currently in?
36. What is the name of the lights on the TPWS panel?
37. Do the signals in this area have three aspects or four aspects?
38. What is the current speed of your train?
39. What would you do if you saw a person trackside waving both hands above their head?
40. What is the current main reservoir pipe reading?
41. If you have just passed a signal showing a single yellow, what would you do if the next signal showed no aspect?
42. What was the name of the last station you passed?
43. What would you do if your mobile phone were to ring when you are driving a train and approaching a station?
44. What time is currently displayed on your Mitrac screen?
45. What aspect was the last signal you passed displaying?
46. What would you do if you came across smoke from a bonfire blowing across the track and partially obscuring a signal?
47. What notch is your power brake control currently in?
48. If you are working a train with a conductor and receive 2 bells on the bell buzzer system while stopped at a station, what does this mean?
49. Do the signals in this area have three aspects or four aspects?
50. What is the current speed of your train?
51. What would you do if you were travelling at 30mph and saw animals on the line in the distance?
52. What is the current brake pipe reading?
53. When you are departing a station and you receive a single long tone on the bell buzzer system, what does this mean?

54. Name one person who can authorise you to pass a signal which is showing a danger aspect.

55. What is the maximum speed for passing over points or crossings when you are passing a signal at danger with the signaller's authority?

Appendix C Demographic questionnaire

To be completed by Researcher Participant Number: _____________	Time slot: _________	Date: _____/_____/_____

Background information

1.	**What was your age at your last birthday?**		
2.	**Are you Male or Female? (please tick)**	Male	Female
4.	**How old were you when you became certified to drive a train? (leave blank if you are a trainee)**		
5.	**How many years train driving experience do you have?**		
6.	**Have you ever driven on this train simulator before? (please tick)**	YES	NO

7. Please tick the category which best describes your role at Southern:			
Current mainline driver	Trainer of other drivers	Trainee train driver	Other (please specify)
8. If you selected "trainer of other drivers", when did you cease regular train driving? (please write in the space provided)			

Appendix D Karolinska sleepiness scale (KSS)

1 Extremely Alert
2
3 Alert
4
5 Neither alert nor sleepy
6
7 Sleepy - but no difficulty in keeping awake
8
9 Extremely sleepy - fighting sleep

Appendix E NASA TLX

To be completed by Researcher		
Participant Number: _____________	Drive number: __________	Date: _____/_____/_____
NASA TLX		

Your experience of the last drive — NASA TLX

For the following questions please think about the drive you just completed an "X" along each scale at the point that best indicates your experience.

Some of the scales may seem strange at first glance. If you're not confident that you have understood the descriptions of the scales, please do not hesitate to ask an experimenter for further clarification

1 **Mental Demand:** How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the drive easy or demanding, simple or complex, exacting or forgiving?

Low |_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_| High

2 **Physical demand:** How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc.)? Was the drive easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Low |_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_| High

3 **Temporal demand**: How much time pressure did you feel due to the rate or pace at which the drive occurred? Was the pace leisurely or rapid and frantic?

Low |_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_| High

4 **Performance**: How successful do you think you were in accomplishing the goals of the drive? How satisfied were you with your performance in accomplishing these goals?

Low |_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_| High

5 **Effort**: How hard did you have to work (mentally and physically) to accomplish your level of performance?

Low |_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_| High

6 **Frustration**: How discouraged, stressed, irritated, and annoyed verses gratified, relaxed, contented, and complacent did you feel during your drive?

Low |_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_| High

Appendix F Stress Arousal Checklist (SACL)

To be completed by Researcher		
Participant Number: _____________	Drive number: _________	Date: _____/_____/_____

How did you feel during the drive?

Please read the following adjectives and circle the most appropriate response for each one, based on how you felt during the drive you have just completed. Thus,

- if a word definitely describes how you have felt then circle DEF
- if a word describes slightly how you have felt then circle SL
- if you do not understand or cannot decide circle UN
- if a word definitely does not describe how you felt circle NOT

Make sure you circle one of the responses for each word. Please complete this questionnaire quickly, your first reaction is best.

1	Worried	DEF	SL	UN	NOT
2	Peaceful	DEF	SL	UN	NOT
3	Active	DEF	SL	UN	NOT
4	Drowsy	DEF	SL	UN	NOT
5	Fearful	DEF	SL	UN	NOT
6	Calm	DEF	SL	UN	NOT
7	Activated	DEF	SL	UN	NOT
8	Stimulated	DEF	SL	UN	NOT
9	Contented	DEF	SL	UN	NOT
10	Dejected	DEF	SL	UN	NOT
11	Up-tight	DEF	SL	UN	NOT
12	Vigorous	DEF	SL	UN	NOT
13	Alert	DEF	SL	UN	NOT
14	Idle	DEF	SL	UN	NOT
15	Passive	DEF	SL	UN	NOT
16	Cheerful	DEF	SL	UN	NOT
17	Uneasy	DEF	SL	UN	NOT
18	Jittery	DEF	SL	UN	NOT
19	Pleasant	DEF	SL	UN	NOT
20	Comfortable	DEF	SL	UN	NOT
21	Lively	DEF	SL	UN	NOT
22	Restful	DEF	SL	UN	NOT
23	Tired	DEF	SL	UN	NOT
24	Sleepy	DEF	SL	UN	NOT
25	Bothered	DEF	SL	UN	NOT
26	Apprehensive	DEF	SL	UN	NOT
27	Tense	DEF	SL	UN	NOT
28	Nervous	DEF	SL	UN	NOT
29	Sluggish	DEF	SL	UN	NOT
30	Aroused	DEF	SL	UN	NOT
31	Distressed	DEF	SL	UN	NOT
32	Relaxed	DEF	SL	UN	NOT
33	Somnolent	DEF	SL	UN	NOT
34	Energetic	DEF	SL	UN	NOT

Appendix G Post drive interview

Mitigation interview/questionnaire

Train Drivers' Attitudes and Perceptions Towards

The Task of Answering Questions

(Interactive Secondary Task)

When this interview is administered: After 2nd drive with mitigation on 2nd drive

This interview is designed to gather some feedback about your attitudes and perceptions towards being asked task-related questions whilst driving. This feedback will really help us to incorporate the views of the drivers and to gain an understanding into how you feel about the questions. Participation in this interview is entirely voluntary and your answers will remain completely anonymous. Are you happy for me to record this interview so that I can make detailed notes later? You will not be identifiable from the recordings. This is quite a brief interview and should only last about 10 minutes. Please answer honestly and give as much detail as you can. Also, please try to provide answers based on your own thoughts rather than how you think other drivers might feel.

We'll start off with a few simple questions.

1. How did you feel about being asked questions whilst you were driving?
    a. Why?

____

____

____

2. How did you feel about the questions themselves?
    a. The frequency
    b. The difficulty
    c. The terminology
    d. Appropriateness

____

____

____

3. How engaged were you with the task of answering the questions?
    a. Did you hardly think about the questions/answers
    b. Did you think hard about the questions/answers

____

____

OK. So now I'd like to find out about how answering the questions may have affected you.

4. Do you think the task of answering the questions affected your **driving performance** during the trial? Can you provide details?
    a. If yes, how did it change tour performance?
    b. How much did it change your performance?
    c. Ask about specific tasks, e.g. monitoring the environment, maintaining constant speed.

Overall, what impact do you think the questions had on your driving? Positive impact/negative/no impact

____

____

____

____

____

____

____

5. Did you think task of answering the questions affected **you,** and how you felt during the drive?

    POSSIBLE PROMPTS:
    a. Mood
    b. Boredom
    c. Fatigue
    d. Motivation
    e. Effort
    f. Attention/alertness
    g. Interest
    h. Enjoyment
    i. Overall effect – positive/negative/neutral?

____

____

____

____

____

OK. So these questions have concentrated on how you felt about answering the questions in the simulator, now we can talk about how this may relate to driving in real trains.

6. How would you feel about answering questions when you are driving a **real train**?
    a. Benefits
    b. Drawbacks
    c. How do you think other drivers would feel about?

7. Do you have any other comments about answering questions whilst driving a train?

The task of answering the questions was designed to help reduce the negative effects of low workload on driving performance. The theory is that, under conditions in which the driver has little to do, a secondary task of answering questions will slightly increase the driver's workload. This increase in workload should relate to an increase in performance (show the performance/workload curve).

8. Do you think this could be an effective mitigation for low workload?

    a. Why?

9. Do you think the task of answering questions could be used in real train driving scenarios to prevent low workload?

    b. Why?
    c. What are the problems with using this task?
    d. What are the benefits of using this task?

10. Do you think there are any improvements we could make to this mitigation?

    e. To make it more effective?
    f. To make it more acceptable?

____

____

____

____

1. Do you have any ideas about what would make an effective mitigation to prevent low workload in train drivers?

____

____

____

Closing

Thank you for your answers. They will be really helpful. Just to summarise, in general, you felt that the task of answering questions was ________________________________ and it had an overall effect on you and your performance. You also felt that answering questions in a real train was ____.

Is there anything else you would like to add?

____

____

That concludes the interview and the final part of the study. The study is now finished. Thank you for participating in the study.