The development and validity of an objective indicator of fatigue for frontline safety critical workers

Michael Scott Evans

Thesis submitted to Cardiff University, School of Psychology in partial fulfilment of the requirements for the award of the Degree of Doctor of Philosophy (PhD)

Centre for Occupational and Health Psychology
School of Psychology, Cardiff University, UK

2019

Summary

The aim of this thesis was to develop and validate an objective indicator of fatigue for frontline safety critical workers. This thesis was carried out in partnership with Arriva Trains Wales (ATW). The rationale for developing an alternative objective indicator of fatigue stems from the fact that the current biomathematical model of fatigue used at ATW was found to be an ineffective predictor of train driver's fatigue levels. In addition, observations from inside the cabin identified that noise, environmental temperature, incomplete train improvements, and cabin working conditions were also major issues that could contribute towards safety incidents when fatigued. As a result, clear evidence was found that an alternative objective indicator of fatigue was needed to support the fatigue risk management system (FRMS) at ATW. In a controlled laboratory setting, the 10-minute psychomotor vigilance task (10-min PVT) has become the widely accepted 'gold standard' tool for assessing the impact of sleep deprivation and fatigue on human cognitive neurobehavioral performance for monitoring temporal changes in attention. Therefore, several studies were carried out to replicate and validate an alternative online mobile version of the 10-min PVT i.e., online 10-min m-PVT, a shorter version i.e., online 5-min m-PVT as well as developing an offline iOS mobile app version i.e., offline 10-min m-PVT. Findings from these studies identified that the online 10-min m-PVT using the time-of-day and time-on-task paradigm was sufficiently sensitive in detecting levels of fatigue, while the online 5-min m-PVT was able to provide an objective indicator of simulated workload fatigue. The offline 10-min m-PVT was also found to be sensitive at detecting levels of fatigue for train drivers in their operational setting. Further research is now needed to investigate whether a shorter offline 5-min m-PVT could still be sensitive enough at detecting levels of fatigue for frontline safety critical workers.

Table of Contents

List of Figures

List of Tables

Acknowledgement

In loving memory of Irwan Ahmad, Christine Noble and Sasha.

During my time at Cardiff University I have been fortunate enough to meet some of the most remarkable people. I would like to firstly take this opportunity to thank Professor Andrew Smith and Prof Rob Honey for their continuous support, inspiration and guidance throughout my doctoral research at Cardiff University. I want to also thank Prof Tom Freeman for taking the time to provide invaluable insights. I am also extremely grateful to Gwilym Bowen, for not just helping me throughout my data collection process at Arriva Trains Wales (ATW), but also for taking the time to mentor me. I cannot be thankful enough. Thank you very much Gwilym for everything you have done for me. I would also like to thank Daniel Harborne who has been an incredible friend, inspiration, motivation and collaborator. The direction of this thesis would not have been possible without you! I also want to thank Rhiannon Doller, Sue Anzeze, Matthew John and the whole team at the occupational health and safety division as well as everyone from the resource management team office for always being there to help me and engaging in valuable discussions throughout my three years. Without your continuous support and guidance, this thesis would not have been possible.

I would like to take this opportunity to also thank Gareth Jones, Dave Rees, Richard Wylie, Andrew Gazzard, Louise Gerrish, Sally Rees, Greg Wilcox, James Regan, and Phil Turner for taking the time to meet with me and help me further understand ATW operational structure and systems. A special thank you goes to my four train driving instructors Michael Joyce, Nigel Tylor, Steven Willis and Dewi Lewis who kindly offered to let me sit inside the cabin to gain a real first-hand experience when it comes to some of the operational challenges being faced each and every day. From the bottom of my heart, thank you so much for making me feel welcomed and allowing me to be feel like an honorary train driver. I also want to take this opportunity to thank each and every single train driver who kindly took the time to take part in the study. I cannot thank you all enough. Without your participation, passion and openness that together we can work together for a better future, this research would not have been possible. I had been a real pleasure working with the team at ATW. I have had a wonderful time working and being welcomed and taken on as an extended family member within the safety department.

I would like thank Ian Bullock and the board of directors at ATW for their openness and willingness for me to carry out this project. I want to specially thank the ESRC for funding the project in collaboration with ATW. I am also truly indebted and thankful to all the participants who took time out of their busy lives to take part in the research. It was an honour to meet such thoughtful, considerate and friendly individuals.

I also want to take this opportunity to thank Becca Louch for all her continuous dedication, support and inspiration throughout my PhD journey. A huge thank you also goes to Federica Vallone and the whole Vallone family – I love you all! I also want to thank each one of my friends and colleagues for such a wonderful three years: Eman Alharbi, Hasah Alheneidi, Louise Bowen, John Galvin, Jennifer Langer, Kenisha Nelson, Shaffika Nor, Omolaso Omosehin, and Gareth Richards. I want to also thank Kirsty Firman and Kai Thomas for all their help and support. I also want to thank my extended family at the BPS Welsh Branch and PsyPAG for peer-support throughout my PhD journey.

I would also like to take this opportunity to thank all professional service staff at the School of Psychology for taking the time to assist me above and beyond of what would ever be expected. In alphabetical order thank you: Paul Allen, Dr Claudia Calder, David Griffiths, Felicity Harrett, Laura James, Hilmar Jay, Mark Jones, Lisa Kennedy, Judy McPherson, Rhian Newman, Hannah Parker, Laura Parsons, Sarah Parsons, Elaine Rees, Lawrence Roche, Alison Roberts, Lindsay Roberts, Scot Welsh, John Woods, and Lorraine Woods.

I would like to also thank one of the most incredible hardworking individuals I have ever met – Dr Yi-Hsin Lin (Jessica), who not only has always believed in me in each step of my journey, but also provided me with continuous support throughout some of the most challenging moments of my life. I would also like to thank my family: Anne Marie Evans, Sandie Cash, Tyne Evans, Kent Evans, Elaine Evans, Leon Ward, and Chris Ward for always being there to support me and providing much needed encouragement throughout my two PhD journeys.

Publication and impact in Thesis

Section of Chapter 3 and Chapter 7 have been presented in the following publication. Moreover, due to the nature of this research, impact has also been generated by presenting as an invited keynote speaker.

1.1: List of publications

1. **Evans, M. S.** (2019). *Fatigue in frontline safety critical workers: Investigating the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator as an objective indicator of fatigue*. Unpublished internal document, Health and Safety Authority (HSA), Republic of Ireland.

2. **Evans M. S.**, Harborne, D., & Smith, A. P. (2019). Exploring an objective indicator of fatigue: A modified mobile version of the Psychomotor Vigilance Task (m-PVT). In: L. Longo & M. C. Leva (Eds.), *Human Mental Workload: Models and Applications. H-WORKLOAD 2018. Communications in Computer and Information Science 1012* (pp. 49-71). Springer, Cham. doi: 10.1007/978-3-030-14273-5_4

3. **Evans, M. S.** (2018). *Fatigue in safety critical roles*. Unpublished internal document, Transport for Wales Rail Services (TfWRS).

4. **Evans, M. S.** (2018). *Fatigue in safety critical roles*. Unpublished internal document, Arriva Trains Wales (ATW).

5. **Evans, M. S.**, & Smith, P. A. (2018). Analyzing safety incidents using the Fatigue Risk Index calculator as an indicator of fatigue within a UK train franchise. *World Academy of Science, Engineering and Technology, International Science Index, Medical and Health Sciences*, 12, 3042.

6. **Evans, M. S.** (2018). *Fatigue in train drivers*. Unpublished internal document, Arriva Trains Wales (ATW).

7. **Evans, M. S.** (2017). *Arriva Trains Wales' (ATW) Fatigue Risk Index (FRI) parameters at each depot*. Unpublished internal document, Deutsche Bahn AG, Germany.

8. **Evans, M. S.** (2017). *Safety incidents: An examination of the Fatigue Risk Index (FRI) calculator*. Unpublished internal document, Arriva Trains Wales (ATW).

1.2: List of keynote speaker invitations

1. **Evans, M. S.** (2018, December). *Fatigue in safety critical roles*. Invited by Dr Chiara Leva (Lecturer in Health and Safety) to present to the School of Food Science and Environmental Health, Dublin Institute of Technology. Dublin, Republic of Ireland.

2. **Evans, M. S.** (2018, August). *Impact of fatigue in train drivers*. Invited by Mr Martyn Brennan (Director of Operations) to present to the Executive Board at Arriva Trains Wales (ATW). Cardiff, UK.

3. **Evans, M. S.** (2018, August). *Developing an objective indicator of fatigue for train drivers*. Invited by Mr Dan Basacik (Human Factor Specialist, Rail Safety and Standards Board) to present at the Rail Safety and Standards Board (RSSB). London, UK.

4. **Evans, M. S.** (2018, May). *An overview of fatigue in train drivers at Arriva Trains Wales (ATW)*. Invited by Mr Ken Skates AM (Minister for Economy and Transport) to present to the Welsh Government Transport for Wales (TfW). Cardiff, UK.

5. **Evans, M. S.** (2018, May). *An overview of fatigue in train drivers: Key findings from the last three years of research*. Invited by Mr David Rees (Head of Drivers) to present to Arriva Trains Wales (ATW) Driver Manager Team Meeting. Cardiff, UK.

6. **Evans, M. S.** (2018, April). *Using a mobile app to explore workload fatigue in train drivers*. Invited by Professor Lorraine Whitmarsh (Director of Environmental Psychology) to present to the Transport Futures Research Network. Cardiff, UK.

7. **Evans, M. S.** (2018, February). *Fatigue in train drivers at Arriva Trains Wales (ATW)*. Invited by Dr Claire Dickinson to present to the Office of Rail and Road (ORR). Manchester, UK.

8. **Evans, M. S.** (2018, February). *Fatigue in train drivers*. Invited to present by members of two train drivers' unions; the Associated Society of Locomotive Steam Enginemen and Firemen (ASLEF) and the National Union of Rail, Maritime and Transport (RMT) to present to Arriva Trains Wales (ATW) Company Council Meeting. Cardiff, UK.

9. **Evans, M. S.** (2017, August). *Using mobile technology to objectively measure fatigue*. Invited by Mr Jim Taylour (Head of Design and Wellbeing) to present to orangebox. Newport, UK.

1.3: Podium conference presentations

1. **Evans, M. S.**, Harborne, D., & Smith, A. P. (2018, September). Exploring an objective measure of fatigue: A mobile version of the Psychomotor Vigilance Task (m-PVT). Paper presented at the H-Workload 2018 Conference. Amsterdam.

2. **Evans, M. S.**, & Smith, P. A. (2018, July). How objective is the Fatigue Risk Index (FRI) calculator at identifying safety incidents that fatigue could have been a contributing factor? Paper presented at the 33rd PsyPAG Annual Conference 2018. University of Huddersfield, UK.

3. **Evans, M. S.**, & Smith, P. A. (2018, March). *Analyzing safety incidents using the Fatigue Risk Index calculator as an indicator of fatigue within a UK rail franchise*. Paper presented at the ICOHS 2018: 20th International Conference on Occupational Health and Safety. Miami, US.

4. **Evans, M. S.**, & Smith, A. P. (2017, October). *Can fatigue be objectively measured using an iPhone?* Paper presented at the School of Psychology Thursday Talk Series. Cardiff, UK.

5. **Evans, M. S.** (2017, August). *How fatigued are you feeling right now?* Paper presented at the Nuffield Research Placement: Speaker Series. Cardiff University, UK.

6. **Evans, M. S.**, & Smith, A. P. (2017, July). *Can fatigue be objectively measured using mobile devices?* Paper presented at the 32nd PsyPAG Annual Conference 2017. Northumbria University, UK.

1.4: Poster conference presentations

1. **Evans, M. S.**, & Smith, A. P. (2018, May). *Developing an objective indicator of fatigue for train drivers*. Poster presented at the Centre of Artificial Intelligence, Robotics and Human Machine Systems event. Cardiff, UK.

2. **Evans, M. S.**, & Smith, A. P. (2018, May). *Using the Fatigue Risk Index (FRI) calculator as an indicator of fatigue for train drivers*. Poster presented at the Centre of Artificial Intelligence, Robotics and Human Machine Systems event. Cardiff, UK.

3. **Evans, M. S.**, & Smith, A. P. (2018, May). *An objective indicator of fatigue: An online mobile version of the Psychomotor Vigilance Task (m-PVT)*. Poster presented at the Centre of Artificial Intelligence, Robotics and Human Machine Systems event. Cardiff, UK.

4. **Evans, M. S.**, & Smith, A. P. (2018, January). *Developing an objective indicator of fatigue for train drivers*. Poster presented at the Human Factors Excellence at Cardiff University (HuFEx). Cardiff, UK.

5. **Evans, M. S.**, & Smith, A. P. (2018, January). *Using the Fatigue Risk Index (FRI) calculator as an indicator of fatigue for train drivers*. Poster presented at the Human Factors Excellence at Cardiff University (HuFEx). Cardiff, UK.

6. **Evans, M. S.**, & Smith, A. P. (2018, January). *An objective indicator of fatigue: An online mobile version of the Psychomotor Vigilance Task (m-PVT)*. Poster presented at the Human Factors Excellence at Cardiff University (HuFEx). Cardiff, UK.

7. Curtis, E., & **Evans, M. S.** (2017, October). *Incident management using mobile technology in the transport industry*. Poster presented at the Nuffield Research Placement 2017 Celebration Event, Techniquest, Cardiff, UK.

8. Jarvisto, I., & **Evans, M. S.** (2017, October). *Creating flat cartoon images for a semantic processing mobile app*. Poster presented at the Nuffield Research Placement 2017 Celebration Event, Techniquest, Cardiff, UK.

9. Jones, J. S. M., & **Evans, M. S.** (2017, October). *The effects of overtime on train drivers relating to the number of incidents*. Poster presented at the Nuffield Research Placement 2017 Celebration Event, Techniquest, Cardiff, UK.

10. Khan, S. Z., & **Evans, M. S.** (2017, October). *An investigation into train drivers working overtime and how it impacts fatigue*. Poster presented at the Nuffield Research Placement 2017 Celebration Event, Techniquest, Cardiff, UK.

11. Morris, O., & **Evans, M. S.** (2017, October). *Objective measures of fatigue using mobile technology*. Poster presented at the Nuffield Research Placement 2017 Celebration Event, Techniquest, Cardiff, UK.

12. **Evans, M. S.**, & Smith, A. P. (2016, November). *A mobile version of the Psychomotor Vigilance Task (m-PVT) implemented on Apple's iPhone 6s Plus*. Poster presented at the School of Psychology PG Conference 2016. Cardiff, UK.

13. Chiganze, T., & **Evans, M. S.** (2016, November). *The link between train crew fatigue and safety incidents*. Poster presented at the Nuffield Research Placement 2016 Celebration Event, Techniquest, Cardiff, UK.

14. Dhariwal, J., & **Evans, M. S.** (2016, November). *An investigation of whether train crew fatigue index scores are related to incidents*. Poster presented at the Nuffield Research Placement 2016 Celebration Event, Techniquest, Cardiff, UK.

15. Rivera, S., & **Evans, M. S.** (2016, November). *Train drivers and fatigue*. Poster presented at the Nuffield Research Placement 2016 Celebration Event, Techniquest, Cardiff, UK.

16. **Evans, M. S.**, & Smith, A. P. (2016, October). *Objective measures of fatigue: A modified mobile version of the Psychomotor Vigilance Task (m-PVT)*. Poster presented at the WISERD Annual Conference 2016, Swansea, UK.

17. **Evans, M. S.**, & Smith, A. P. (2016, May). *Investigating objective measures of fatigue using an alternative version of the Psychomotor Vigilance Task (PVT) implemented in mobile devices*. Poster presented at the ESRC PGR Conference 2016. Liverpool, UK.

18. **Evans, M. S.**, & Smith, A. P. (2015, November). *Objective measures of fatigue in Train Crew*. Poster presented at the School of Psychology PG Conference 2015. Cardiff, UK.

Train operating company (TOC) context

In December 2003, Arriva Trains Wales (ATW) under parent company Arriva was awarded the contract to operate the Wales and Borders Franchise for 15 years until 14th October 2018. Four companies entered bids to the Welsh Government to operate the new Wales and Borders rail service and the South Wales Metro, with a contract to operate from 4th June 2018 – 14th October 2033. These four companies were; Arriva Trains Wales (ATW), Abellio Rail Cymru, KeolisAmey and MTR Corp (Cymru) Ltd. In October 2017, Arriva withdrew their bid to operate the Wales and Borders rail service and the South Wales Metro. In February 2018, Abellio Rail Cymru also withdrew their bid to operate the Wales and Borders rail service and the South Wales Metro due to their inability to meet the tender requirements outlined by TfW after their development partner Carillion entered liquidation. As a result, KeolisAmey and MTR Corp (Cymru) Ltd were the two remaining bidders. In May 2018, The Welsh Government awarded the French-Spanish joint venture KeolisAmey the contract to operate the Wales and Borders rail service and the South Wales Metro, which will be overseen by Transport for Wales (TfW). As of the 14th October 2018, KeolisAmey is currently operating the Wales and Borders rail service and the South Wales Metro under the name Transport for Wales Rail Services (TfWRS), with the aim to transform rail travel over the next 15 years. Since the present PhD thesis was carried out between October 2015 – September 2018 in partnership with Arriva Trains Wales (ATW), all references throughout this PhD thesis will therefore solely address and acknowledge Arriva Trains Wales (ATW) as the Wales and Borders Franchise and not the current Wales and Borders rail service and the South Wales Metro – Transport for Wales Rail Services (i.e., KeolisAmey).

Abbreviations

1/RT	Response Speed
5-min m-PVT	5-minute mobile Psychomotor Vigilance Task
10-min m-PVT	10-minute mobile Psychomotor Vigilance Task
ADD	Adjusted Diagram Duration
AFI	Adjusted Fatigue Index
ANOVA	Analysis of Variance
ARI	Adjusted Risk Index
ASAP	Aviation Safety Action Program
ASLEF	Associated Society of Locomotive Steam Enginemen and Firemen
ATO	Automatic Train Operation
ATW	Arriva Trains Wales
AWS	Automatic Warning System
BAC	Blood Alcohol Content
BMI	Body Mass Index
BMM	Biomathematical Model
CCMTA	Canadian Council of Motor Transport Administrator
CCT	Camel and Cactus test
CED	Crewe Depot

CFD	Cardiff Depot
CFS	Chronic Fatigue Syndrome
CHD	Chester Depot
CHS	Centre for Human Science
CND	Carmarthen Depot
CSD	Cardiff Shed Depot
CSS	Cascading Style Sheets
CVD	Cardiff Valleys Depot
DAS	Driver Advisory Systems
DD	Diagram Duration
DERA	Defence Evaluation and Research Agency
DMU	Diesel Multiple Unit
DOC	Drivers Company Council
DRA	Driver's Reminder Appliance
DSD	Driver Safety Device
DVT	Driving Van Trailer
ECG	Electrocardiogram
EDA	Electrodermal Activity
EEG	Electroencephalogram

EMS	Experimental Management System
ERTMS	European Rail Traffic Management System
FAA	Federal Aviation Administration
FAID	Fatigue Audit Interdyne
FI	Fatigue Index
FRA	Federal Railroad Administration
FRI	Fatigue Risk Index
FRMS	Fatigue Risk Management System
GBR	Great Barrier Reef
GDPR	General Data Protection Regulation
GSM–R	Global System for Mobile Communications – Railway
GSR	Galvanic Skin Response
HDD	Holyhead Depot
HSE	Health and Safety Executive
HR	Heart Rate
HTML	Hypertext Markup Language
IFCS	Incident Factor Classification System
IUMI	International Union of Marine Insurance
KSS	Karolinska Sleepiness Score

LED	Light-Emitting Diode
LJD	Llandudno Junction Depot
LTP	Long-Term Planning
MND	Machynlleth Depot
MRB	Main Rest Break
m-PVT	mobile Psychomotor Vigilance Task
NASA-MATB	National Aeronautics and Space Administration Multi-Attribute Task Battery
NICE	National Institute for Health and Care Excellence
NTS	Non-Technical Skills
OGE	Oil and Gas Extraction
ORR	Office of Rail Regulation / Office of Rail and Road
OTMR	On-Train Monitoring Recorder
PAO	Passenger as Ordered
PPE	Personal Protective Equipment
PTSD	Post-Traumatic Stress Disorder
PVT	Psychomotor Vigilance Task
PVT-B	Psychomotor Vigilance Test-Brief
PWD	Pwllheli Depot
RAIB	Rail Accident Investigation Branch

REC	Railway Emergency Group Call
RHD	Rhymney Depot
RI	Risk Index
RMT	National Union of Rail, Maritime and Transport Workers
ROGS	Railways and Other Guided Transport Systems (Safety) Regulations 2006
RoSPA	Royal Society for the Prevention of Accidents
RS	Rail Services
RSSB	Rail Safety and Standards Board
RT	Reaction Time
RTAs	Road Traffic Accidents
SAM	Search and Memory Task
SB	Standby
SFA	Simon Folkard Associates
SMIS	Safety Management Information System
SP	Spare
SPAD	Signal Passed at Danger
SPSS	Statistical Package for the Social Sciences
SQD	Shrewsbury Depot
SRA	Strategic Rail Authority

SRB	Short Rest Break
SRT	Simple Reaction Time
STP	Short-Term Planning
STUD	Safety, Training and Update Day
TB	Team Briefing
TDLCR	Train Driving Licences and Certificates Regulations 2010
TfW	Transport for Wales
TfWRS	Transport for Wales Rail Services
TOC	Train Operating Company
TPWS	Train Protection & Warning System
TTD	Treherbert Depot
UMP	Unified Model of Performance
WHO	World Health Organization

Organisation of Thesis

1.1: Aim of Thesis

The overarching aim of this thesis was to develop and validate an objective indicator of fatigue for frontline safety critical workers.

1.2: Objectives of Chapters

1.2.1: Chapter 1: General introduction

Chapter 1 provides the rationale for studying occupational fatigue in frontline safety crucial workers. This chapter also provides a brief discussion of the research context as well as the objectives being addressed in the thesis.

1.2.2: Chapter 2: Theoretical framework and rationale of the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator used at Arriva Trains Wales (ATW)

Chapter 2 provides the theoretical framework of the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator used at Arriva Trains Wales (ATW) as part of their Fatigue Risk Management System (FRMS) strategy to manage fatigue levels in frontline safety critical workers.

1.2.3: Chapter 3: The Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator used at Arriva Trains Wales (ATW)

Chapter 3 investigates the effectiveness of the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator used at Arriva Trains Wales (ATW) as part of their Fatigue Risk Management System (FRMS) for monitoring and managing safety incidents in which fatigue could have been a contributing factor. Chapter 3 set out to answer two fundamental questions:

1. The aim of the first study was to investigate whether the present biomathematical model (BMM) for assessing train drivers' level of fatigue at Arriva Trains Wales

(ATW) i.e., the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator related to the number of safety incidents in which fatigue could have been a contributing factor.

2. The aim of the second study was to investigate whether restricting train drivers from working on their assigned rest days based on a naturally occurring intervention reduced the number of safety incidents in which fatigue could have been a contributing factor.

1.2.4: Chapter 4: Pre-unit mobilisation procedures and checks as well as in-cab observations

Chapter 4 identifies some of the external environmental factors that could contribute towards safety incidents when fatigued through ethnographic research based on extensive in-cab observations.

1.2.5: Chapter 5: Developing and validating an alternative online objective mobile indicator of fatigue

Chapter 5 investigates the use of the time-of-day and time-on-task effect to replicate and validate whether the alternative online mobile version of the 'gold standard' 10-minute Psychomotor Vigilance Task (10-min PVT) i.e., online 10-minute mobile Psychomotor Vigilance Task (online 10-min m-PVT) could be used to provide an objective indicator of fatigue for frontline safety critical workers.

1.2.6: Chapter 6: Investigating a shorter mobile version of the online 10-min m-PVT i.e., online 5-min m-PVT as an objective indicator of simulated workload fatigue

Chapter 6 investigates whether a shorter mobile version of the online 10-minute mobile Psychomotor Vigilance Task (online 10-min m-PVT) i.e., online 5-min m-PVT could be used to provide an objective indicator of simulated workload fatigue.

1.2.7: Chapter 7: Developing and validating an offline iOS mobile app version of the online 10-min m-PVT i.e., offline 10-min m-PVT for frontline safety critical workers

Chapter 7 outlines the development and validation of an alternative offline iOS mobile app version of the online 10-minute mobile Psychomotor Vigilance Task (online 10-min m-PVT) i.e., offline 10-minute mobile Psychomotor Vigilance Task (offline 10-min m-PVT) to detect levels of fatigue for frontline safety critical workers.

1.2.8: Chapter 8: General Discussion

Chapter 8 provides a general discussion of the findings in relation to the objectives of the thesis. This chapter summarises the overarching theoretical and practical implications of the thesis, limitations of the thesis, practical recommendations, recommendations for future research, the future of train driving, and concluding remarks.

Chapter 1: Introduction

1.1: Research objectives

The originality and contribution of this thesis lies in three main domains; Firstly, identifying whether the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator used by Arriva Trains Wales (ATW) is an effective method for monitoring and reducing safety incidents in which fatigue could have been a contributing factor; Secondly, to better understand some of the issues that could contribute towards safety incidents when fatigued through ethnographic research based on extensive in-cab observations; Thirdly, to develop, validate, and investigate an alternative objective indicator of fatigue in frontline safety critical workers e.g., train drivers, hospital staff, emergency services, and law enforcement that can be used to strengthen and further complement the current Fatigue Risk Management System (FRMS) at ATW. Therefore, all three main domains were addressed by answering the following research objectives (see Table 1).

<table>
<tr><td>Objective 1: To investigate the effectiveness of the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator used at Arriva Trains Wales' (ATW) as part of their Fatigue Risk Management System (FRMS) for monitoring and managing safety incidents in which fatigue could have been a contributing factor. As a result, the first objective set out to answer two fundamental questions:
1. The aim of the first study is to investigate whether the present biomathematical model (BMM) for assessing train drivers' level of fatigue at Arriva Trains Wales (ATW) i.e., the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator related to the number of safety incidents in which fatigue could have been a contributing factor.
2. The aim of the second study was to investigate whether restricting train drivers from working on their assigned rest days based on a naturally occurring intervention reduced the number of safety incidents in which fatigue could have been a contributing factor.</td></tr>
<tr><td>Objective 2: To identify some of the external environmental factors that could contribute towards safety incidents when fatigued, through ethnographic research based on extensive in-cab observations.</td></tr>
<tr><td>Objective 3: To use the time-of-day and time-on-task effect to replicate and validate whether the alternative online mobile version of the 'gold standard' 10-minute Psychomotor Vigilance Task (10-min PVT) i.e., online 10-minute mobile Psychomotor Vigilance Task (online 10-min m-PVT) could be used to provide an objective indicator of fatigue in frontline safety critical workers.</td></tr>
</table>

Objective 4: To investigate whether a shorter mobile version of the online 10-minute mobile Psychomotor Vigilance Task (online 10-min m-PVT) i.e., online 5-min m-PVT could be used to provide an objective indicator of simulated workload fatigue.
Objective 5: To develop and validate an alternative offline iOS mobile app version of the online 10-minute mobile Psychomotor Vigilance Task (online 10-min m-PVT) i.e., offline 10-minute mobile Psychomotor Vigilance Task (offline 10-min m-PVT) to detect levels of fatigue for frontline safety critical workers.

Table 1: Research objectives

1.2: General introduction

According to McOrmond (2004), the modern context way in which people work has dramatically changed in the last few decades from a manufacturing-based to a service-based economy. Therefore, due to societal demands, almost all large organisations now trade and operate throughout the 24-hour clock (Ritson & Charlton, 2006). As a result, today's labour workforces are increasingly working non-standardised shift patterns (Beers, 2000; Krausz, Sagie, & Bidermann, 2000; Presser, 2003). Consequently, the 09:00 – 17:00 working model may no longer be the standardised shift pattern (Jamal, 2004; Pisarski, Lawrence, Bohle, & Brook, 2008). Patrick and Gilbert (1896) were some of the first scientists to empirically determine some of the physiological and mental effects of enforced abstinence from sleep. In their work, three subjects were kept awake for 90 hours, and during this period performed a series of tests of functions such as; reaction time, discrimination-time, motor ability, memory and attention at 6 hours' intervals. In this study the researchers were able to demonstrate the ill effects of remaining awake for such a prolonged duration of time.

1.3: Definition of fatigue

According to Kroemer and Grandjean (1997), fatigue is often defined as the decline in mental and/or physical performance that results from prolonged exertion, lack of quality sleep or disruption of the internal body clock. The Office of Rail and Road (ORR) defines fatigue as '*a state of perceived weariness that can result from prolonged working, heavy* *workload, insufficient rest or inadequate sleep*' (RSSB, 2012: 12). Bowler and Gibson (2015: 4) further added that fatigue can be '*a feeling of extreme tiredness and being unable to perform work effectively.*' However, there are other factors which may impact on, or contribute to, higher levels of fatigue which should also be taken into consideration. Both these physical and mental factors include individual, environmental, and work-related factors (see Figure 1; RSSB, 2012). Furthermore, there are also personal factors that may also contribute towards higher levels of fatigue, such as home life commitments and conflicts (HSE, 2006). Earlier research by Krueger (1989) identified that the consequences of fatigue may include reduced cognitive performance, reduced alertness and increased levels of sleepiness.

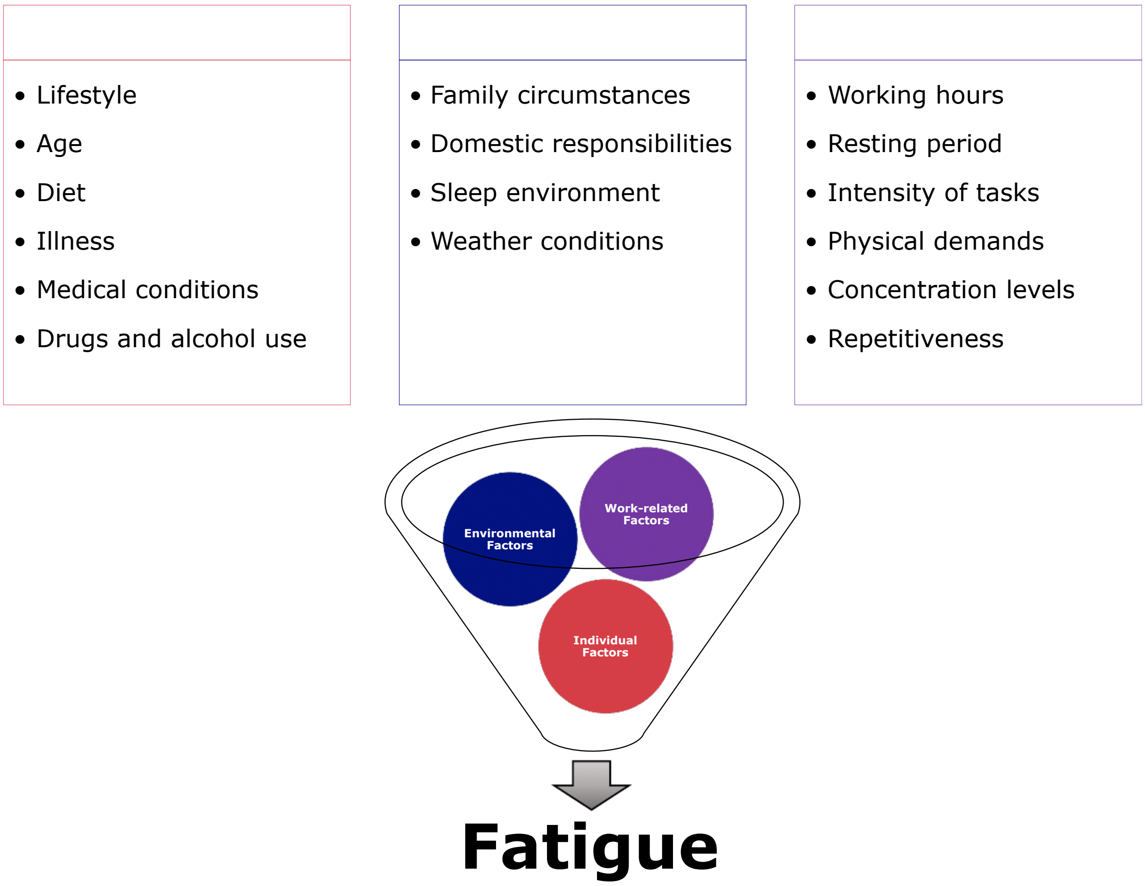


Figure 1: Fatigue factors (adapted from RSSB, 2012)

1.4: The fatigue process

Since both physical and mental fatigue can occur simultaneously, a simple way of conceptualising fatigue would be to consider a fatigue process. According to Smolarek and Soliwoda (2014), fatigue is a process, which begins with the actual risk factors for fatigue, and moves on to the subjective perceptions of fatigue, and finally ends with the consequences of fatigue (see Figure 2). Smith, Allen and Wadsworth (2015), added that the fatigue process, which starts with risk factors of fatigue includes long working hours and sleep loss. This is then followed by consideration of the individual's perceptions of fatigue. The process finally ends with the outcomes of fatigue, which include changes in mood, inefficient performance, accidents, and injuries.

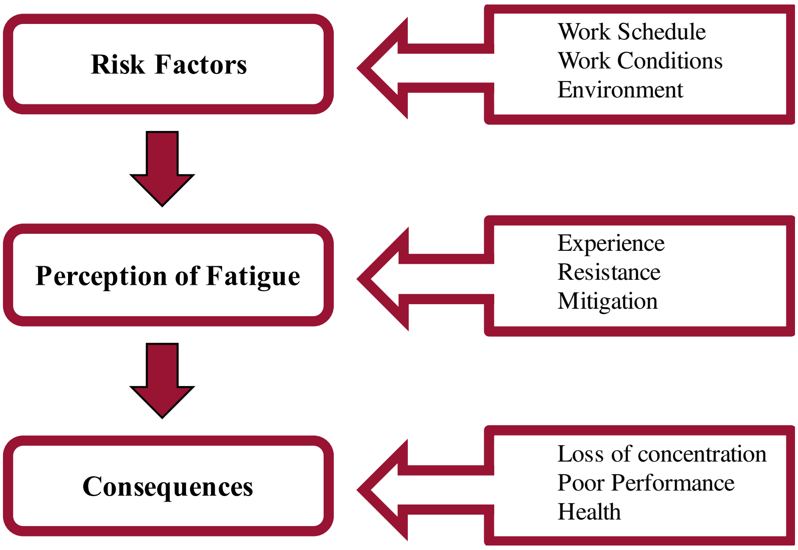


Figure 2: Fatigue Process Model (adapted from Smolarek & Soliwoda, 2014)

1.5: Benchmarking fatigue

Research carried out by Dawson and Reid (1997) found that the effects of moderate fatigue on performance are similar to moderate alcohol intoxication. In other words, continued wakefulness of around 17 hours resulted in a decrease in cognitive performance, which was the equivalent to deficit observed performance at a blood alcohol content (BAC) of ~.05 per cent. In addition, after 24 hours of continued wakefulness, performance decreased to a level equivalent to a BAC of .1 per cent. However, a BAC will also depend on multiple individual factors, such as weight, age, sex, percentage of alcohol in the beverage, and the rate of drinking (see Winek & Esposito, 1985, for review). As a result, Brick (2006) developed a series of mathematical calculations, which aimed to standardise and integrate individual factors with BAC. To put these figures into perspective, the National Highway Traffic Safety Administration (2010) identified that in 2007 in the United States alone, there were 55,681 drivers involved in fatal crashes that resulted in 41,059 deaths, in which 22 per cent of these drivers were legally intoxicated (BAC $\geq$ .08 per cent).

1.6: Occupational fatigue

1.6.1: Prevalence

It is estimated that approximately 20 per cent of the employed population reported symptoms of feeling fatigued at work (Bültmann, Kant, Kasl, Beurskens, & van den Brandt, 2002). However, these findings may not be representative of all working population sectors and industries. Other studies have reported prevalence rates of fatigue ranging from 7 per cent to 45 per cent (Lewis & Wessely, 1992). According to Lewis and Wessely (1992), these discrepancies could be due to the methodology used in the study, as well as the cut off points that were selected in those studies in. As a result, fatigue has become a serious problem due to an ever-increasing workplace workload, long duty working hours, disruptions on circadian rhythms, pressure from social and societal demands as well as sleep deprivation (Caldwell, Caldwell, Thompson, & Lieberman, 2019; Sadeghniiat-Haghighi & Yazdi, 2015). Therefore, the impact of fatigue at work can clearly have significant consequences and implications on employees personal, environmental, and industrial everyday experience (Smith, Allen, & Wadsworth, 2015). However, there are other factors that could cause increased fatigue levels. For example, sleep deprivation (Lo *et al*., 2012; Simpson *et al*., 2016), shiftwork (Gorlova *et al*., 2019), and chronic fatigue (Caldwell *et al*., 2019).

1.6.2: Sleep deprivation

Lo *et al.* (2012) identified that partial sleep deprivation i.e., less than seven hours of sleep per night (Shockey & Wheaton, 2017) across several days impairs both cognitive performance and heath, whether physical health issues or mental health issues (Dahl & Lewin, 2002). Simpson *et al.* (2016) outlined that this is due to the fact that it is extremely difficult for anyone to be able to adapt to prolonged partial sleep deprivation. Tononi and Cirelli (2006) argued that impaired cognitive performance is the most rapidly occurring consequence of the effect of sleep deprivation. In addition, it has been found that sleep deprivation results in fatigue (Killgore, 2010; Legault, Clement, Kenny, Hardcastle, & Keller, 2017) as well as various studies having identified that sleep deprivation leads to chronic fatigue (Ahsberg *et al.*, 2000; Kerin & Aguirre, 2005; Muecke, 2005).

1.6.3: Chronic and acute fatigue

Workers are always confronted with several unforeseeable physical, cognitive, and emotional demands in the workplace (Querstret, Cropley, Kruger, & Heron, 2016). These demands do consume valuable physical and mental resources (Meijman, Mulder, & van Dormolen, 1992), which can result in acute and chronic fatigue over prolonged periods of consecutive exposure (van der Ploeg & Kleber, 2003). According to Fukuda *et al.* (1994), chronic fatigue syndrome (CFS) is a clinically defined condition that is characterised by severe disabling fatigue that includes a combination of self-reported symptoms such as impairments in concentration as well as short-term memory loss, sleep deprivation, and musculoskeletal pain.

CFS has been linked to cognitive deficits (Capuron *et al.*, 2006; Costigan, Elliott, McDonald, & Newton, 2010; Hou *et al.*, 2014; Santamarina-Perez, Eiroa-Orosa, Rodriguez-Urrutia, Qureshi, & Alegre, 2014; Thomas & Smith, 2009) as well as an increased susceptibility to acute fatigue (Smith *et al.*, 1999). However, Robinson *et at.*

(2019) and Beaumont *et al*. (2012) have identified that cognitive deficits in CFS may not be as broad as previously suggested (e.g., Thomas & Smith, 2009) and may simply be restricted to slowing in basic processing speed. At present, there are no diagnostic tools beyond symptom recognition and no curative treatments (NICE, 2007; NICE 2018a; NICE 2018b; NICE 2018c; NICE 2018d). However, with better understanding of CFS and technological advancements, researchers are now exploring new approaches to potentially locate biomarkers for the diagnosis of CFS (Xu *et al*., 2019). For the time being, research carried out by Thomas and Smith (2006) has identified that antidepressant therapy was able to reduce patients' CFS, even after three years. In addition, rehabilitation courses have also proven to be successful using self-directed management techniques of CFS (Harrison, Smith, & Sykes, 2002). Furthermore, employees working in medium or high-risk professions, e.g., train drivers, hospital staff, emergency services, and law enforcement, etc. are often confronted with acute stressors or critical incidents and thus may be involved in life threatening situations (Querstret, Cropley, Kruger, & Heron, 2016). In addition, Mitchell and Dyregrov (1993) have shown that acute stressors or critical incidents may lead to serious mental disturbances, and in most cases post-traumatic stress disorder (PTSD). In contrast, Martin, Hebert, Ledoux, Gaudreault and Laberge (2012) identified that there is a relationship between work-related fatigue and chronotype to sleep. Therefore, van Dongen (2006) argues that individual differences in tolerances may have a biological chronotype basis (see Hittle & Gillespie, 2018, for review), when it comes to adapting to fatigue.

1.6.4: Chronotypes

Wittmann, Dinich, Merrow and Roenneberg (2006) defined chronotype as the individual variations of sleep/wake times, which are primary influenced by environmental light, genetics, and human development stages (Wittmann *et al*., 2006). According to Hittle and Gillespie (2018) approximately 60 per cent of the population will fall into the intermediate stage. However, Adan, Archer, Hidalgo, Di Milia, Natale and Randler (2012) outline that chronotypes can range from early birds (i.e., morning chronotype) to night owls (i.e., evening chronotype), which can significantly impact on an individual's ability to adapt to shift work (Hittle & Gillespie, 2018).

Hittle and Gillespie (2018) suggest that reducing employees' need to work night shifts would be extremely beneficial. However, for safety critical workers such as train drivers, hospital staff, emergency services and law enforcement, the ability to reduce shiftwork is not always possible due to the operational need or nature of the safety critical service. As a result, Hittle and Gillespie (2018) argue that there is a clear need for the identification of shift workers chronotype as a means to develop a tool that could help predict disease development.

1.6.5: Shiftwork

It has been identified that shiftwork leads to disruptions in the natural circadian rhythm (Loudoun & Allan, 2008), which contributes towards both psychological and physiological symptoms as well as impacting productivity, increasing the likelihood of being involved in an accident, and making errors more likely (Banks & Dinges 2007; Folkard & Tucker 2003; Rajaratnam & Arendt 2001). Therefore, the most commonly reported issue with shiftwork has been the decrease in the quality and quantity of sleep (Conway *et al*., 2008; Costa & Sartori, 2007; Fletcher & Dawson, 2001; Yong, Li, & Calvert, 2017). This is further supported by numerous studies which have also identified that shiftwork decreases the quality and quantity of sleep (Äkerstedt, Fredlund, Gillberg & Jansson, 2002; Conway *et al*., 2008; Charles, *et al*., 2007; Chang, 2018; Costa & Sartori, 2007; Fletcher & Dawson, 2001), as the sleeping rhythm is well correlated with the circadian rhythm (Lac & Chamoux, 2004). However, sleep quality and quantity are not the only implications of shiftwork (Reid *et al*., 2018). Sasaki *et al*. (2018) identified that poor sleep quality was associated with an increase in health issues among shift workers.

Shiftwork has also been extensively linked to a range of chronic health problems (Khan, Duan, Yao, & Hou, 2018; Matheson, O'Brien, & Reid, 2014; Sasaki *et al*., 2018; Yong, Li, & Calvert, 2017), such as sleep disorders (e.g., Flo *et al*., 2012; Gorlova *et al*., 2019; Zhang Sun, Li, & Tao, 2016), cancer (e.g., Davis, Mirick, & Stevens, 2001; Hansen, 2001; Savvidis & Koutsilieris, 2012; Walasa *et al*., 2018), obesity (e.g., Fonken *et al*., 2010; Liu *et al.,* 2018; Supriyanto, Soemarko, & Prihartono, 2018), and cardiovascular disease (e.g., Barger *et al*., 2017; Ha & Park, 2005; Peter, Alfredsson, Knutsson, Siegrist, & Westerholm, 1999). Kelley, Feltman and Curry (2018) identified that inconsistent shiftwork, insufficient rest, and poor sleep quality are factors that contribute towards fatigue and performance degradation, as it has been well documented that prolonged working hours and displaced shiftwork results in physiological fatigue, mental fatigue (e.g., cognitive performance deficits and errors) and an increase in safety accidents (Mallis, Mejdal, Nguyen, & Dinges, 2004).

The link between shiftwork and safety is far better established within the transport industry, when compared to other industrial sectors (Åkerstedt & Wright, 2009). This is due to the fact that driving a mode of transport that carries passengers e.g., trains, buses, and coaches, demands continuous attention and any lapse in attention can result in drivers being involved in safety critical incidents (Philip & Akerstedt, 2006). However, the implications and consequences of shiftwork go far beyond the physiological and psychological health issues as well as increased safety incidents. Yong, Li and Calvert (2017) argue that shiftwork also results in shift workers experiencing difficulties in performing certain activities of daily living in general, such as the ability to concentrate, working on hobbies and/or taking care of financial affairs (Yong, Li, & Calvert, 2017). Nevertheless, Lian *et al*. (2016) state that establishing effective coping strategies can significantly reduce the impact of shiftwork. Åkerstedt and Landström (1998) found that shift workers utilise various countermeasures as well as behavioural coping strategies to best manage the ill effects of fatigue in order to be able to maintain both safety and performance in the workplace.

Gorlova *et al*. (2019) acknowledge that recovery sleep is a vital ingredient for shift workers to be able to recover from fatigue. Zee and Goldstein (2010) recommend that good sleep practices as well as the application of circadian principles for shift workers could significantly improve sleep quality, alertness, performance, and safety. However, there are other coping strategies that shift workers can use to improve alertness and performance. Walsh, Muehlbach and Schweitzer (1995) found that in a laboratory setting, caffeine was a positive countermeasure of shiftwork-related fatigue and sleep disturbances. However, caffeine has been shown to improve cognitive performance and enhance wakefulness and mood in shift workers (Haskell, Kennedy, Wesnes, & Scholey, 2005; Seidl, Peyrl, Nicham, & Hauser, 2000; Wyatt, Cajochen, Ritz-De Cecco, Czeisler, & Dijk, 2004).

1.6.6: Caffeine and shiftwork

Caffeine (trimethylxanthine) is a purine alkaloid that occurs naturally in coffee beans (Higdon & Frei, 2006) and has become one of the most commonly consumed beverages in the world, accounting for 75 per cent of the regular soft drink consumption (Toci, Farah, Pezza, & Pezza, 2016). However, drinking coffee has often been discouraged (Ciaramelli, Palmioli, & Airoldi, 2019) as increased caffeine consumption has been found to be associated with greater sleep deprivation, psychological distress, abdominal pain and weight gain (Centofanti *et al*., 2018). In addition, poor sleep quality is also related to an increase in caffeine consumption (Dorrian, Baulk, & Dawson, 2011; Hsieh *et al*., 2011). However, Booker, Magee, Rajaratnam, Sletten and Howard (2018) argued that perhaps the reason for the association between poor sleep quality and increased caffeine consumption might be due to shift workers trying to remain alert, and not that the caffeine itself causes poor sleep quality unless caffeine was consumed within close proximity to trying to fall asleep.

Smith (2005) identified that higher levels of caffeine consumption resulted in significantly higher alertness levels over the working day as well as significantly reduced reaction times. In addition, an association between caffeine consumption and fewer cognitive failures and accidents at work was also found (Smith, 2005). Snel and Lorist (2011) argue that caffeine is a popular means to enhance various aspects of cognitive performance. Dekker, Paley, Popkin and Tepas (1993) found that train drivers reported higher caffeine consumption rates than permanent shift factory workers, which resulted in poorer sleep quality, increased negative moods, decreased positive moods during work and during days of rest (i.e., time off). However, Nawrot *et al.* (2003) argued that for the healthy adult population, moderate daily caffeine intake was not associated with several adverse health effects e.g., cardiovascular disease, mood changes, and increased cancer diagnoses. Moreover, Smith, Whitney, Thomas, Perry and Brockman (1997) stated that in stressful situations, low levels of caffeine do not increase physiological and behavioural observed changes. Despite these inconsistencies in the literature, Richards, Stayton, Wells, Parikh and Laurin (2018) identified that caffeine consumption was a well-established coping strategy among shift workers before starting their shift, and both caffeine consumption and napping were found to be the two most ubiquitous fatigue countermeasures utilised by shift workers (Dorrian *et al.*, 2011; Knauth & Hornberger 2003; Roth, 2012; Zee & Goldstein, 2010). As a result, caffeine consumption as well as napping were found to be effective coping strategies for mitigating fatigue levels (Centofanti *et al.*, 2018) as reviews of the literature agree that caffeine consumption improves reaction time (Einother & Giesbrecht, 2013; Nehlig, 2010; Smith, 2002). Smith

(2013) suggested that there could also be possible biological mechanisms underlying the effects of caffeine on cognitive performance. Smith, Sutherland and Christopher (2005) found that caffeine consumption improves performance as well as positive moods. In addition, Smith *et al*. (2013) and Snel and Lorist (2011) have identified that ingestion of caffeine improves 'lower' cognitive functions e.g., simple reaction time and the ability to encode new information more efficiently. However, caffeine consumption effects on 'higher' cognitive functions e.g., decision-making and problem-solving are often debated (Kosslyn & Smith, 2001). Novak and Auvil-Novak (1996) identified that early shift caffeine consumption improved night shift performance. Akerstedt and Ficca (1997) argue that caffeine consumption is commonly used to improve alertness during work. However, caffeine consumption has been shown to be more effective at particular ingestion times. Walsh, Muehlbach and Schweitzer (1995) identified that a single dose of caffeine at the start of a night shift was more alerting than several divided doses throughout the night shift. McLellan, Caldwell and Lieberman (2016) highlight that under normal day-to-day circumstances, there is evidence to suggest that caffeine consumption is modulated until a self-perceived optimal peak level of arousal and cognitive performance is achieved (Harvanko, Derbyshire, Schreiber, & Grant, 2015). Therefore, Brice and Smith (2001) argued that when performance and alertness levels are depleting or significantly low, caffeine is an effective coping strategy for sustaining attention, alertness and positive mood. It is also important to point out that the effects of caffeine have been integrated into some biomathematical model (BMM), such as the unified model of performance (UMP) (Ramakrishnan *et al*., 2016). However, caffeine consumption counteracts reductions in the turnover of central noradrenaline, which may underlie the beneficial effects of caffeine that are seen in low alertness levels (Smith, Brice, Nash, Rich, & Nutt, 2003).

1.6.7: Noradrenaline and attention lapses

Smith and Nutt (1996) have demonstrated that noradrenaline is important in maintaining attention. Greene, Bellgrove, Gill and Robertson (2009) argue that sustained attention is the capacity to keep oneself alert or the ability to remain focused on a particular task for a prolonged duration. Posner and Peterson's (1990) theory of attention outlines that the alerting network responsible for arousal and vigilance relies strongly on the actions of the neurotransmitter noradrenaline. As a result, elevated noradrenaline levels have been shown to enhance responses by reducing response latencies (Bouret & Sara, 2002; Lecas, 2004) and by minimising response thresholds (Ciombor, Ennis, & Shipley, 1999; Waterhouse, Azizi, Burne, & Woodward, 1990; Waterhouse *et al.*, 1988). Therefore, noradrenaline seems to improve attention by firstly narrowing the focus of attention and secondly by blocking out the effect of distractors (Robbins, 1984; Smith, Wilson, Glue, & Nutt, 1992).

1.7: Fatigue in the transport industry

Williamson *et al.* (2011) identified that the consequences of fatigue results in an increase in safety incidents. Therefore, occupational fatigue has been widely studied in various frontline safety critical sectors of the transport industry, including aviation (e.g., Bennett, 2003; Caldwell & Gilreath, 2002), road transport industry (e.g., Dawson, Searle, & Paterson, 2014), truck drivers (e.g., Williamson, Feyer, & Friswell, 1996), seafarers (e.g., Lützhöft, Dahlgren, Kircher, Thorslund, & Gillberg, 2010; Smith, Allen, & Wadsworth, 2006; Wadsworth, Allen, McNamara, & Smith, 2008; Wadsworth, Allen, Wellens, McNamara, & Smith, 2006), and the railway industry (e.g., Darwent, Lamond, & Dawson, 2008; Dorrian, Baulk, & Dawson, 2011; Dorrian, Hussey, & Dawson, 2007; Dorrian, Roach, Fletcher, & Dawson, 2007; Jay, Dawson, Ferguson, & Lamond, 2008; Ku & Smith, 2010; Sussman & Coplen, 2000). In addition, this has also been found in law enforcement officers (e.g., Hursh, *et al.*, 2004; Sanquist, Raby, Forsythe, & Carvalhais, 1997); emergency services (e.g., Paley & Tepas, 1994). However, the true extent of the problem of fatigue within transportation remains unknown (Noy *et al*., 2011). Nevertheless, Rosekind *et al.* (1996) point out that the task of managing fatigue in an operational setting e.g., the transport industry (maritime, aviation, railway, etc.) should be a shared responsibility across all stakeholders e.g., among individuals, companies, federal agencies, scientists, industry organisations as well as the public. Smith (2016) found that poor driving behaviour, driving when fatigued, and risk taking were the main predictors of road traffic accidents (RTAs).

1.7.1: Road fatigue and road traffic accidents (RTAs)

Road traffic accidents (RTAs) represents a significant threat to human lives all around the world (Li, Yamamoto, & Zhang, 2018), with an estimated 1.25 million fatalities on the road each year and millions more sustaining serious life altering injuries (WHO, 2015). In addition, driving when fatigued was identified as one of the four most serious and risky driving-related behaviours, especially in fatal RTAs (Fernandes, Hatfield, & Job, 2010). However, fatigue involvement in fatal RTAs crashes vary considerably. For example, it has been found that fatigue was a significant contributing factor in ~20 per cent of all fatal RTAs in Canada (CCMTA, 2010) as well as the United Kingdom (RoSPA, 2017). These estimates were even higher in Australia, being between 20 and 30 per cent (ATC, 2018). As a result, driver fatigue is now considered to be a major contributor of all RTAs by approximately 15 – 30 per cent (Anund, Ihlstrom, Fors, Kecklund, & Filtness, 2016; Connor *et al*., 2002; Williamson *et al*. 2011). Conversly, Williamson *et al.* (2011) found no supporting evidence between circadian-related fatigue influences and performance or safety outcomes. Nevertheless, Smith and Allen (2013) compared fatigue in lorry drivers and seafarers and found similar self-reported fatigue levels. In addition, Anund *et al.* (2016) examined sleepiness in city bus drivers and found that 19 per cent of city bus drivers had to fight to stay awake. Furthermore, Anund *et al*. (2016) found that severe sleepiness correlated with fatigue related safety risks e.g., near crashes. Akhtar and Utne (2014) argue that a significant percentage of maritime accidents are due to fatigue, and that the causes of fatigue at sea have been categorised in terms of both physical and mental workload aspects (Besikci *et al*., 2016)

1.7.2: Maritime fatigue

Besikci, Tavacioglu and Arslan (2016) have identified that the maritime industry has always had a long record of accidents. The International Union of Marine Insurance (IUMI) have argued that human factors are a contributing element in the rise of maritime transport accidents (Nilsson, Garling, & Lutzhoft, 2009), while human error remains the main factor for a large proportion of maritime accidents (Darbra & Casal, 2004; O'Neil, 2003; Toffoli, Lefevre, Bitner-Gregersen, & Monbaliu, 2005; Tzannatos, 2010). Smith (2008) outlined that offshore workers experienced poorer sleep quality and significant sleep deprivation.

1.7.2.1: Offshore fatigue (oil installations)

Riethmeister, Brouwer, van der Klink and Bultmann (2016) found that 73 per cent of offshore workers reported prolonged fatigue due to shiftwork. Mehta *et al.* (2017) argued that offshore workers are exposed to intensive shift patterns as well as long work durations, which lead to high levels of fatigue. Mathisen and Bergh (2016) identified that duration of employment as well as overtime increased action errors and rule violation of offshore workers. In addition, Riethmeister *et al.* (2016) found that 46 per cent of offshore workers were overweight i.e., body mass index (BMI) ≥ 25, while 21 per cent of the offshore workers were obese i.e., BMI ≥ 30. These BMI thresholds were within the reference methods for evaluating total body fat levels corresponding to the BMI thresholds (Gallagher *et al*., 2000), i.e., underweight (< 18.5), overweight (≥ 25), and obesity (≥ 30). Furthermore, fatigue does not exclusively impair performance or reduce safety during shiftwork (Smith, Allen, & Wadsworth, 2006) but can also impact safety outside of work. Mason, Retzer, Hill and Lincoln (2015) identified that approximately 40 per cent of fatalities from the oil and gas extraction (OGE) industry occurred as workers return home from their offshore shifts. Therefore, Mette, Garrido, Harth, Preisser and Mache (2018) state that interventions should be introduced to better promote effective coping strategies. Conversely, Salyga and Kusleikaite (2011) identified that in seafarers the duration of work had a dramatic effect on the prognosis of fatigue symptoms, with an incredible 87 per cent stating that they experienced fatigue-related symptoms.

1.7.2.2: Seafarers' fatigue

Smith and Lane (2001) outlined that approximately 25 per cent of seafarers had experienced fatigue during their watch, and out of those approximately 50 per cent had expressed that fatigue reduced their ability to evaluate dangerous situations, as well as making it more difficult to predict potential accidents, while others simply reported that they had fallen asleep. Research carried out by Smith, Allen and Wadsworth (2006) on seafarers found that the consequences of fatigue are not simply the immediate visible evidence of reduced safety and impaired performance, but also in the decreased well-being and increased risk of mental health problems, which are particularly prevalent in the development of chronic diseases among seafarers (Jepsen, Zhao, & van Leeuwen, 2015).

Looking more closely at some examples of accidents in the maritime sector as a result of fatigue, one can clearly see how serious fatigue impacts different situations. On the 3rd April 2010, *Shen Neng 1* a Chinese-registered coal ship grounded on Douglas Shoal, a section of the Great Barrier Reef (GBR) in Australia. An independent investigation by the Australian Transport Safety Bureau highlighted that there were a number of factors which contributed to the accident. However, the grounding ultimately happened as a result of the chief mate failing to adjust the ship's course due to not monitoring the ship's position. The investigators found that the chief mate had only slept for 2 hours and 30 minutes in the preceding 38 hours and 30 minutes before the grounding. The investigation concluded that the chief mate had made a succession of errors, which were probably brought about by his lack of sleep (Australian Transport Safety Bureau, 2011). Research carried out by Smith, Allen and Wadsworth (2006) found that approximately 16 per cent of fishermen had been involved in a fatigue related incident. More alarming, 60 per cent of fishermen said that their personal safety had been at risk because of fatigue at work and that 44 per cent had worked to the point of collapse. However, it is important to note that this study was based on a small-scale survey (n = 81) of fishermen who completed the fishing fatigue questionnaire. Moreover, the study did not go further to determine the nature of the incident beyond the single-item questions or investigate what active role the fishermen played in the incident. However, fatigue symptoms are not limited to maritime settings alone and thus have also been extensively explored within the aviation industry (Driskell & Mullen, 2005; Goode, 2003; Kelley *et al*., 2018; Lee & Kim, 2018; Morris, Wiedbusch, & Gunzelmann, 2018; Neville, Bisson, French, Boll, & Storm, 1994).

1.7.3: Aviation fatigue

It is estimated that approximately 70 per cent of all aviation accidents can be attributed to human factors (Rudari, Johnson, Geske, & Sperlak, 2016; Yen, Hsu, Yang, & Ho, 2009). Neville *et al*. (1994) found that lengthy work periods, reduced sleep, and shiftwork were associated with higher fatigue levels as well as an association between fatigue and pilot error. Conversely, Pellegrino and Marqueze (2019) identified that self-perception of insufficient sleep increases perception of fatigue. In other words, if a shift worker personally believed (s)he did not get sufficient sleep, then the shift worker is more likely to express symptoms of fatigue. Therefore, within the aviation industry, shiftwork, insufficient rest, and poor sleep quality were factors contributing to fatigue and performance degradation (Kelley *et al*., 2018). Despite serious fatigue related safety concerns in the aviation industry, there are no tests for fatigue either before or after a safety accident (Goode, 2003).

An examination of Aviation Safety Action Program (ASAP) reports from the aviation industry identified that fatigue was a contributing factor to safety incidents whereby aircraft operation violations were the most cited consequences of fatigue as well as fatigue being almost twice as likely to be reported as a secondary rather than primary contributing factor (Morris, Wiedbusch, & Gunzelmann, 2018). Research examining fatigue in commercial aircraft pilots identified a discernible pattern of increased probability of a safety accident occurring as shift duration increased (Goode, 2003).

Lee and Kim (2018) identified that psychological and physiological decline, as well as rest failings, were risk factors that affected pilot fatigue, which can be even further broken down into seven independent variables i.e., flight direction, crew scheduling, partnership, aircraft environment, job assignment, ethnic difference, and hotel environment. Conversely, there are also concerns that the scheduling of pilot's diagrams can lead to fatigue, which inevitably increases the changes of an aviation accident occurring (Goode, 2003). Nevertheless, Driskell and Mullen (2005) argue that the benefits of napping in the aviation industry could provide an effective fatigue countermeasure. Extensive research has shown that fatigue is also a major concern within the rail industry (Dorrian *at al*., 2011; Dunn & Williamson, 2012; Cotrim *at al*., 2017; Harma, Sallinen, Ranta, Mutanen, & Muller, 2002; Korunka, Kubicek, Prem, & Cvitan, 2012; Tsao, Chang, & Ma, 2017).

1.7.4: Railway fatigue

Fatigue is a frequent problem in the railway industry due to irregular shift schedules (Harma *et al*., 2002). Most notably, it was identified that fatigue caused by extreme overtime by the train driver was a contributory factor in the 1988 Clapham Junction collision, which killed 35 people (RSSB, 2012). Within the last 10 years, the Rail Accident Investigation Branch (RAIB) investigated and concluded that fatigue played a key role in some of the highest profile rail incidents in Great Britain (Bowler & Gibson, 2015). As a result, failure to manage fatigue in the rail industry properly can have disastrous consequences, as evident from the following incidents:

Date	Incident	Location
09.02.2006	Derailment of the freight train	Melton Mowbray
26.04.2008	Freight train collision	Leigh-on-Sea
10.10.2008	Derailment of two locomotives	East Somerset Junction
17.08.2010	Uncontrolled freight train	Between Shap and Tebay

Note: Adapted from Bowler and Gibson (2015)

Data from the Federal Railroad Administration (FRA, 2015) has identified that in 2015 alone, there were a total of 1,874 train accidents, in which human factors accounted for 735 of all train accidents, representing a significant 39 per cent contribution. However, it is important to note that these figures are based on data from only a few train companies that operate in the United States; Amtrak, BNSF Railway Co.; Canadian National – North America; Canadian Pacific Railway Co.; CSX Transportation; Kansas City Southern Railway Co.; Norfolk Southern Corporation; and Union Pacific Railroad Company (FRA, 2015). Therefore, these figures may not be an actual representation of the severity of the situation, and thus may be a relatively conservative estimate of the actual total number of train accidents, including those caused by human factors in the United States. As a result, the understanding of human factors within the rail industry is becoming significantly prominent (Balfe, Sharples, & Wilson, 2018; Madigan, Golightly, & Madders, 2016; Naghiyev, Sharples, Ryan, Coplestone, & Carey, 2017; Schock, Ryan, Wilson, Clarke, & Sharples, 2010).When looking at statistics within the United Kingdom alone, Bowler and Gibson (2015) have examined high risk fatigue-related railway safety incidents and found a total of 246 fatigue-related safety incidents, in which 53.8 per cent were Signal Passed at Danger (SPAD). As a result of devastating incidents, fatigue and loss of sleep have been globally recognised as a major fundamental safety concern in the rail industry (Edkins & Pollock, 1997), with serious costs and implications both socially (Bowler & Gibson, 2015; RSSB, 2012) and economically (e.g., Lauber & Kayten, 1988).

Dorrian *at al.* (2011) investigated the effects of shiftwork on levels of fatigue and found a significant influence of sleep length, wakefulness, work duration, and workload. Cotrim *et al*. (2017) identified that there was a high prevalence of sleepiness among railway control centre staff during the night shift. Similar results were also found from Harma *et al*.'s (2002) study which outlined that self-reported sleepiness levels when compared to day shift were 6 – 14 times higher for night shift and about twice as high in the morning shift. These findings are not too surprising when considering that night shift leads to disruptions in the natural circadian rhythm (Loudoun & Allan, 2008). However, Korunka, Kubicek, Prem and Cvitan (2012) demonstrated that railway controllers' levels of fatigue while on shift were not only influenced by the quality of rest before the shift, but also by the shift onset perceived workload during the shift. However, the respective roles and responsibilities of railway controllers and train drivers are significantly different.

Tsao, Chang and Ma (2017) stated that fatigue levels for train drivers were directly influenced by working overtime and insufficient rest. Dorrian, Roach, Fletcher and Dawson (2007) found that as fatigue levels in train drivers increased, extreme speed violations (i.e., 25 per cent above the track section limit) and penalty brake applications (automatic vigilance systems) were recorded, which are the Australian's equivalent to the United Kingdom's Train Protection Warning Systems (TPWS) and Automatic Warning System (AWS), respectively (see RSSB, 2015a, for review). However, Kazemi, Mazloumi, Saraji and Barideh (2016) identified that there were similar reported fatigue and workload levels in train drivers after completing a short-haul trip (i.e., ~150 miles) and a long-haul trip (i.e., ~560 miles). However, both types of trips differed significantly in terms of shift duration, consecutive driving hours, amount of rest, and also the number of train drivers in the cab. Nevertheless, Darwent, Lamond and Dawson (2008) found that train drivers incurred a significant cumulative loss of sleep throughout the duration of operating the unit (i.e., train). However, despite the cumulative loss of sleep, train drivers were able to sustain vigilance performance. Nonetheless, it is important to highlight that only ten train drivers completed Darwent *et al.*'s (2008) study.

Dunn and Williamson (2012) identified that the combination of monotony and low task demands had a serious detrimental effect on performance. In contrast, it was also found that with just minor increases in train driver's cognitive demand, such changes were able to significantly mitigate the adverse monotony-related effects on performance. Nevertheless, research from within the rail industry has once again identified that the ability to integrate the coping strategy of napping before commencing shiftwork has been shown to benefit train crew from the ill effects of fatigue (Darwent, Dawson, Paterson, Roach, & Ferguson, 2015; Jay, Dawson, Ferguson, & Lamond, 2008).

Dorrian, Lamond, Kozuchowski and Dawson (2008) investigated whether a new device, designed to detect lowered states of arousal using electrodermal activity (EDA), would be sufficiently sensitive at detecting sleepiness and fatigue levels on train drivers while on a train driving simulator. In their study it was found that the 10-minutes psychomotor vigilance task (10-min PVT) and subjective measures indicated increase levels of sleepiness and fatigue during sustained wakefulness. However, there has been little research within the rail industry that has investigated fatigue using objective measures. A preliminary review of fatigue among railway staff carried out by Fan and Smith (2018) identified that there were only three studies within the rail industry that have investigated fatigue using objective measures, such as the 10-min psychomotor vigilance task (10-min PVT) (Dorrian, Roach, Fletcher, & Dawson, 2007; Dorrian *et al.*, 2008) as well as the Fatigue Audit Interdyne (FAID) software (Darwent *et al.*, 2015; Dorrian, Hussey, & Dawson, 2007). A comprehensive review of the 10-min PVT can be found in Evans, Harborne and Smith (2019) while an overview of the FAID can be found in Dawson and Fletcher (2001). Smith and Smith (2017a) have outlined that at Arriva Trains Wales (ATW), workload measures in frontline safety critical workers do not address human mental workload.

1.7.5: Collaboration with Arriva Trains Wales (ATW)

Fan and Smith (2017a) identified that Arriva Trains Wales (ATW) staff, high workload levels resulted in higher self-reported levels of fatigue. In contrast, Smith and Smith (2017b) argued that noise levels significantly reduced well-being among ATW staff. However, research carried out by Fan and Smith (2017b) found that ATW staff who perceived that they had high levels of support and control reported better work-life balance and an increased sense of well-being.

Previous research carried out at ATW identified that mental workload and working overtime e.g., during rest days were a contributing factor towards fatigue (Fan & Smith, 2019). In addition, Fan and Smith (2019) also explored three predictors of fatigue; physical, mental, and emotional. In this study it was identified that physical fatigue resulted from shiftwork and insufficient time to recover while working, while mental and emotional fatigue was due to inadequately prepared shift patterns, including poor scheduling as well as working extra shifts before assigned time off. However, 42.6 per cent of the sample surveyed (managers and administration staff) worked typical working hours i.e., 09:00 - 17:00, while the remaining 57.4 per cent (train drivers, train guards and station staff) were shift workers. Kelley *et al.* (2018) identified that inconsistent shiftwork, insufficient rest, and poor sleep quality were factors contributing to fatigue and performance degradation. Therefore, there is a clear need to consider the distinction between what can be concluded from shift workers (i.e., train drivers, train guards and station staff) and non-shift workers (i.e., managers and administration staff), since interpretations of analyses were made on role characteristics that are prominent on shift workers and less likely on non-shift workers, e.g., poorly arranged shift and poor timing of shifts, which are predominantly experienced by frontline safety critical workers in the rail industry who are constrained by the operating timetable and operational need of the Train Operating Company (TOC).

Fan and Smith (2017a) investigated the impact of workload and fatigue on performance at ATW staff and found an association between fatigue and perceived levels of stress at work, poor performance, negative work-life balance, and negative well-being. It is important to note that in this study the most common job types were; guards, train drivers, and station workers. Close examination of the data sample revealed that 38.5 per cent of the establishment (i.e., total number of train driver across the franchise) were successfully recruited to take part in this study. In contrast, Smith and Smith (2017a) investigated ATW guards' workload and fatigue levels and found that workload for train guards increased over the working week as well as identifying that workload was correlated with levels of fatigue. These levels of fatigue were in turn associated with higher score from the Health and Safety Executive’s (HSE) Fatigue Risk Index (FRI) calculator. However, further exploration of the data revealed that train guards’ sleep durations did not correlate with levels of workload, effort and fatigue. Nevertheless, fatigue index (FI) scores were predicted by train guards’ sleep duration but not workload or by self-reported fatigue levels.

1.8: Research questions and objectives

The impact of workload on fatigue have been extensively documented in the literature as well as the consequences of fatigue and safety incidents. For frontline safety critical workers, fatigue has been found to be a continuous on-going problem. As a result, the objective biomathematical models (BMMs) of fatigue e.g., the Health and Safety

Executive (HSE) Fatigue Risk Index (FRI) calculator (see Chapter 2, for review) were developed to effectively monitor as well as prevent fatigue related safety incidents. However, determining whether the HSE's FRI calculator at ATW is an effective objective indicator of fatigue have yet to be explored. Moreover, the HSE's FRI calculator is based on pre-defined parameters i.e., short-term fatigue; time of day, shift duration, rest period, and breaks, as well as the build-up in fatigue that is associated with the continual disruption of sleep and the time required to recover from the period of sleep. Therefore, the HSE's FRI calculator fails to take into account other external factors that could contribute towards fatigue beyond the work-related factors (e.g., working hours, resting period, intensity of tasks, commuting time, etc.), such as individual factors (e.g., lifestyle, age, diet, stress levels, medical conditions, drugs and alcohol use, etc.) and environmental factors both external to work (e.g., family circumstances, domestic responsibilities, sleep environment, etc.) and internal to work (e.g., unit condition, track condition, breakroom facilities, napping facilities, weather conditions, etc.). However, studies of the various individual and environmental factors that could intensify fatigue levels, have been widely documented in the literature as discussed previously. Nevertheless, ethnographic research that explores the environmental factors that could contribute towards fatigue from within the cab environment in frontline safety critical train drivers has yet to be carried out. As a result, there is a clear need to identify some of the external environmental factors that could contribute towards safety incidents when fatigued, by the researcher engaging in extensive in-cab observations using ethnography. Further research is also needed to determine whether an alternative objective indicator of fatigue could be developed and validated to support the Fatigue Risk Management System (FRMS) at ATW. Therefore, the thesis is aimed at developing and validating an objective indicator of fatigue for frontline safety critical workers.

1.8.1: Objective 1: Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator used at Arriva Trains Wales (ATW)

As previously emphasised, failing to effectively monitor or manage occupational fatigue in frontline safety critical workers have been shown to result in devastating consequences. Whether these manifest into major safety incidents or health related diseases, occupational fatigue has been shown to dramatically impact not just the frontline safety critical workers but also the social and economic cost. Therefore, it is crucial to investigate the effectiveness of the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator used at Arriva Trains Wales' (ATW) as part of their Fatigue Risk Management System (FRMS) for monitoring and managing safety incidents in which fatigue could have been a contributing factor. As a result, the first objective set out to answer two fundamental questions:

1. The aim of the first study is to investigate whether the present biomathematical model (BMM) for assessing train drivers' level of fatigue at Arriva Trains Wales (ATW) i.e., the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator related to the number of safety incidents in which fatigue could have been a contributing factor.

2. The aim of the second study was to investigate whether restricting train drivers from working on their assigned rest days based on a naturally occurring intervention reduced the number of safety incidents in which fatigue could have been a contributing factor.

1.8.2: Objective 2: Pre-unit mobilisation procedures and checks, as well as in-cab observations

One of the biggest limitations of Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator is that it fails to take into account other fatigue-related factors beyond the work-related factors, such as individual factors and environmental factors both external to work and internal to work. Therefore, exploring the environmental factors that could contribute towards fatigue from within the train cabin would provide valuable insights. As a result, the second objective sets out to identify some of the external environmental factors that could contribute towards safety incidents when fatigued, through ethnographic research based on extensive in-cab observations.

1.8.3: Objective 3: Developing and validating an alternative online objective mobile indicator of fatigue

In a controlled laboratory setting, the human Psychomotor Vigilance Task (PVT) has become the widely accepted 'gold standard' tool for assessing the impact of sleep deprivation on human cognitive neurobehavioral performance for monitoring temporal dynamic changes in attention. The aim of this study is to use the time-of-day and time-on-task effect to replicate and validate whether the alternative online mobile version of the 'gold standard' 10-minute Psychomotor Vigilance Task (10-min PVT) i.e., online 10-minute mobile Psychomotor Vigilance Task (online 10-min m-PVT) and whether it could be used to provide an objective indicator of fatigue in frontline safety critical workers.

1.8.4: Objective 4: Investigating a shorter mobile version of the online 10-min m-PVT i.e., online 5-min m-PVT as an objective indicator of simulated workload fatigue

The aim of this study is to investigate whether a shorter mobile version of the online 10-minute mobile Psychomotor Vigilance Task (online 10-min m-PVT) i.e., online 5-min m-PVT could be used to provide an objective indicator of simulated workload fatigue.

1.8.5: Objective 5: Developing and validating an offline iOS mobile app version of the online 10-min m-PVT i.e., offline 10-min m-PVT for frontline safety critical workers

The aim of this study is to develop and validate an alternative offline iOS mobile app version of the online 10-minute mobile Psychomotor Vigilance Task (online 10-min m-

PVT) i.e., offline 10-minute mobile Psychomotor Vigilance Task (offline 10-min m-PVT) to detect levels of fatigue for frontline safety critical workers.

1.9: Chapter summary

The aim of this chapter was to provide an overview of occupational fatigue within the content of the transport industry as well as to identify any existing gaps in the literature. The next chapter reviews the theoretical framework of the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator used at Arriva Trains Wales (ATW) as part of their Fatigue Risk Management System (FRMS) to monitor and manage fatigue levels in frontline safety critical workers.

Chapter 2: Theoretical framework and rationale of the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator used at Arriva Trains Wales (ATW)

2.1: Overview of chapter

The aim of this chapter is to provide the literature and theoretical framework of the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator used at Arriva Trains Wales (ATW) as part of their Fatigue Risk Management System (FRMS) to monitor and manage fatigue levels in frontline safety critical workers.

2.2: UK rail regulations

After the 1992 General Election, the Prime Minister of the United Kingdom and Leader of the Conservative Party Sir John Major proposed the privatisation of the British Rail. The Railways Act of 1993 was published on 22 January 1993, which dismantled the integrated British Rail (Bowman, 2015), and provided, for the appointment by the Secretary of State for Transport John MacGregor, an officer to be known as 'the Rail Regulator' (Butcher, 2016). The Railways (Safety Critical Work) Regulations 1994 was introduced to address potential risks arising from the fragmentation of the industry following privatisation (Spencer, Robertson, & Folkard, 2006). In 2000, Part IV of The Transport Act 2000 strengthened the power of the Rail Regulator to require the improvement and development of the railway (Butcher, 2016: 4-5) by establishing the Strategic Rail Authority (SRA) with the main purpose:

- '*to promote the use of the railway network for the carriage of passengers and goods;*
- *to secure the development of the railway network; and*
- *to contribute to the development of an integrated system of transport of passengers and goods.*'

However, in 2004, the Rail Regulator was replaced with the Office of Rail Regulation (ORR) under Part 2 of the Railway and Transport Safety Act 2003 to 'streamline the existing structure while still recognising that there will be a continued need for some form of independent economic regulation' (Butcher, 2016: 6). Moreover, in 2005, Part 1 of the Railways Act 2005 was introduced to transfer consumer protection functions of the SRA to the ORR as well as safety functions conferred by the Health and Safety Executive (HSE), despite oppositions worrying that the transfer could potentially increase the safety risk on the railway (Butcher, 2016). In 2015, new road network powers were introduced to the ORR and consequently the Office of Rail Regulation (ORR) became the Office of Rail and Road (retaining the acronym ORR; Butcher, 2016).

At present, the Office of Rail and Road (ORR) is the independent economic and safety regulator for England, Wales, and Scotland (Butcher, 2016) and as from the start of 2017, the economic regulator for Northern Ireland (ORR, 2017a). The train industry funds the ORR, with an estimated annual budget of around £30 million (ORR, 2017a). The budget is used to support the following rail regulatory work (Butcher, 2016: 4):

- regulates Network Rail and High Speed 1 (i.e., the Channel Tunnel Rail Link (CTRL);
- regulates health and safety standards and compliance across the whole rail industry; and
- oversees competition and consumer rights issues.

2.3: Original Health and Safety Executive's (HSE) Fatigue Index (FI)

The original Fatigue Index (FI) was commissioned by the Health and Safety Executive (HSE) and was carried out by the Defence Evaluation and Research Agency (DERA) Centre for Human Science (CHS) to validate and develop an objective biomathematical model (BMM) for assessing risks arising from fatigue that was associated with rotating shift patterns for shift workers involved in safety-critical work (see Rogers, Spencer, & Stone, 1999, for review). In addition, the original FI also served as a tool to provide guidance in support of the Railway (Safety Critical Work) Regulations 1994 (Spencer *et al.*, 2006).

The original FI was developed as an objective BMM of fatigue, which involved the calculations of various components, with the intention to provide the professional rostering team of safety critical workers with an assessment tool for the changes in work pattern (Spencer *et al.*, 2006). In addition, the original FI was also used to determine whether any specific element of the work pattern was likely to contribute towards increased levels of fatigue (Spencer *et al.*, 2006). A final report was produced after two years of research, which highlighted the initial assessment of the index, whereby the strengths and weaknesses of the procedure were identified, followed by data collection (Rogers *et al.*, 1999). Figure 3 outlines the original procedure carried out by CHS.

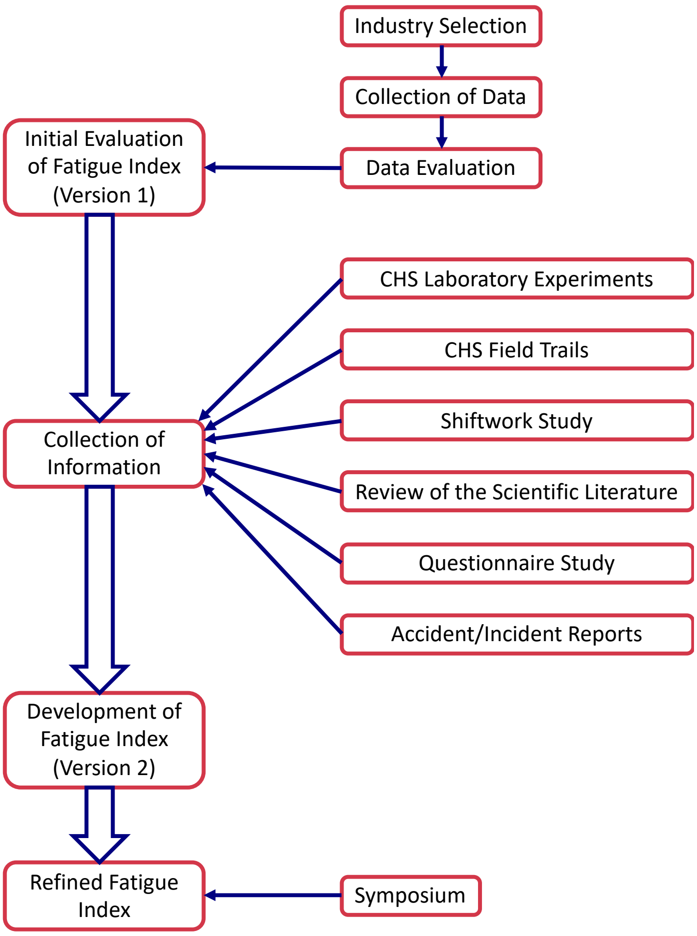


Figure 3: Flow chart of the Centre for Human Science (CHS) work programme, Adapted from (Rogers *et al.*, 1999)

According to Rogers *et al.* (1999), after consultations with representatives from various industries, a revised version was finally produced. The original Fatigue Index (FI) version

2 included five factors (i.e., time of day, shift duration, rest period, breaks and cumulative fatigue), which were known to be related to the build-up of fatigue (see Figure 4). Each factor was assessed independently and added together to give an overall index score for the shift pattern. In terms of how the original FI version 2 was calculated, factors F1, F2, F3, and F4 assessed the short-term fatigue for each individual shift within the roster, while F5 assessed the cumulative fatigue throughout the employee's cumulative working pattern (Rogers *et al.*, 1999). More accurately, the cumulative fatigue component (F5) was developed to represent as simply as possible the build-up of fatigue levels, which were associated with the continual disruption of sleep, together with the time that was required to be able to fully recover from that specific period of disruption (Rogers *et al.*, 1999).

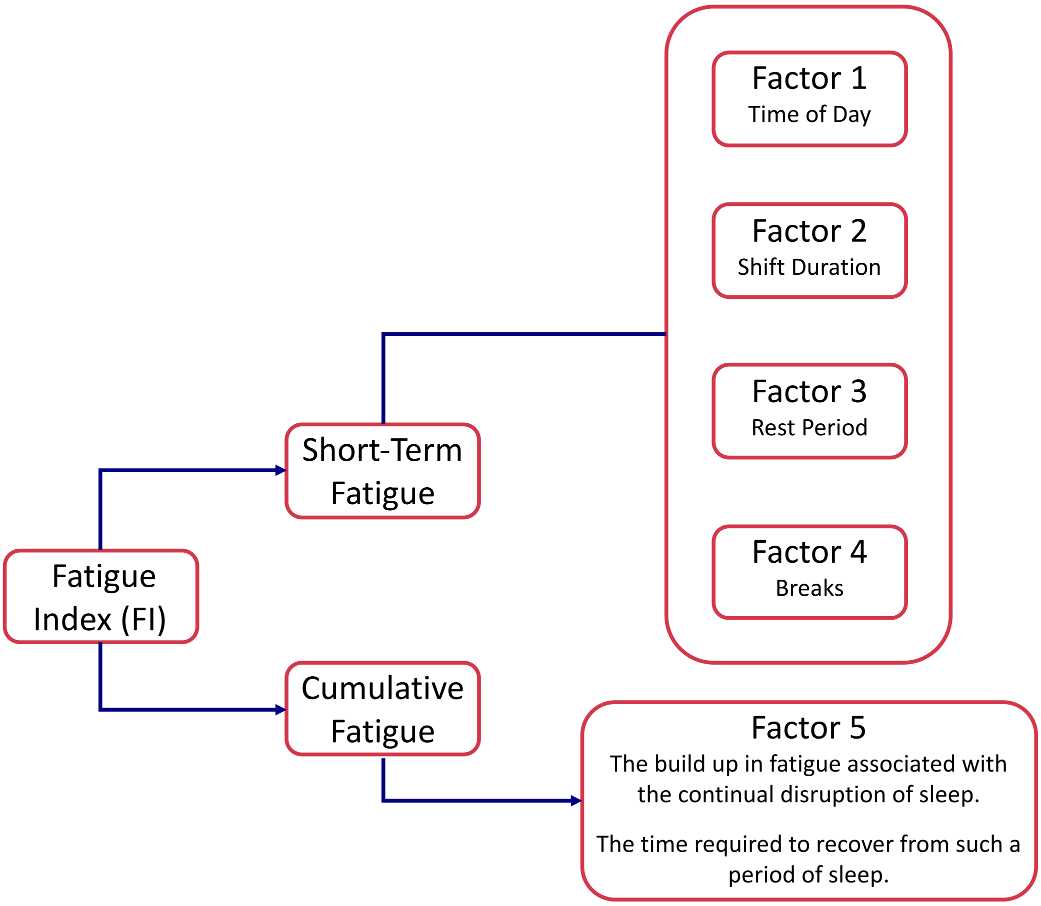


Figure 4: The five factors of the original fatigue index (FI) version 2, Adapted from (Rogers *et al.*, 1999)

According to Spencer *et al*. (2006), the original FI had been widely used by the rail sector as well as being utilised in other areas within the British industrial sector. Kenvyn (2007), outlined that Network Rail had utilised the original FI as a tool by their investigations team to review the working hours of train drivers in the days preceding safety critical incidents. Moreover, the original FI had also been used with train driver interviews to identify whether fatigue was likely to have been a contributing factor towards the safety critical incident (Kenvyn, 2007). However, Spencer *et al*. (2006) pointed out that there were various issues that had been identified with the original fatigue index (FI), which resulted in the requirement for the original FI to be revised. Firstly, it was identified that the original FI was not suitable for long patterns of night shiftwork (Kenvyn, 2007). Secondly, the original FI did not consider factors outside of the workplace that could contribute towards levels of fatigue (Kenvyn, 2007). As a result, the Health and Safety Executive (HSE) once again commissioned an update of the original FI, under contract number 6062 (see Spencer *et al*., 2006, for review).

2.4: Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator

The contract was carried out by QinetiQ, who were previously known as the Defence Evaluation and Research Agency (DERA) in collaboration with Simon Folkard Associates (SFA; Spencer *et al*., 2006). The new biomathematical fatigue index (FI) model is related to the amount of sleep lost that is likely to be associated with the pattern of work and rest (Spencer *et al*., 2006). Therefore, the new FI component was established through two relationships; the relationship between the pattern of duty and sleep and the relationship between the impact of any loss of sleep on levels of fatigue (Spencer *et al*., 2006).

Apart from the subjective rating of workload and attention, the new FI is based on data that are more objective, which allows the various shiftwork arrangements, and rotas to be easily compared (Stanton, Salmon, Jenkins, & Walker, 2010). Therefore, the new biomathematical model (BMM) version of the index included two separate indices to produce two unique scores the fatigue index (FI), and the risk index (RI) (Stanton *et al.*, 2010). Spencer *et al.* (2006) outlined that the main difference between the FI and the RI is related to the effect of time of day, whereby the peak in fatigue occurs around 5am, while the peak in risk occurs close to midnight. As a result, the original FI was updated and renamed to the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator, as it was found that there were indeed considerable differences between the trends in fatigue related to patterns of work and the similar trends in risk (Spencer *et al.*, 2006). As a result, the HSE FRI BMM was designed to assess the level of fatigue experienced by shift workers based on a set of parameters: cumulative effects; duty timing; and job type / breaks (see Figure 5). Therefore, to ensure the HSE FRI calculator is utilised correctly, it is vital to ensure that when an individual's shiftwork pattern is examined that both the FI and RI are reviewed (Spencer *et al.*, 2006).

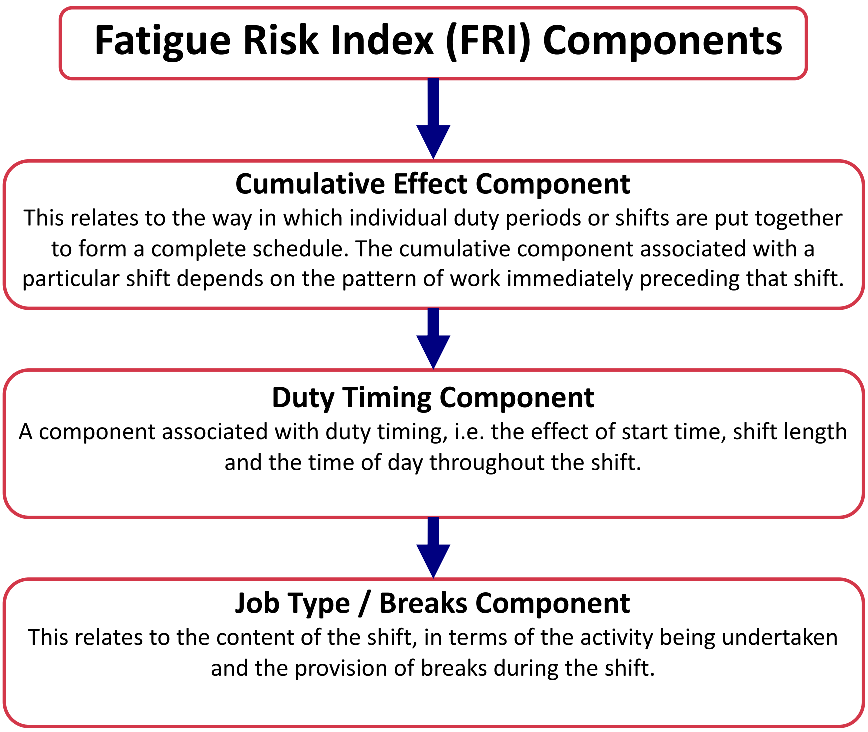


Figure 5: Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator components, adapted from Spencer *et al.* (2006)

2.4.1: Interpretation of the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator biomathematical model (BMM)

When interpreting the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator biomathematical model (BMM) output, i.e., the fatigue index (FI) scores and risk index (RI) scores, it is important that the scores for each shift are considered, rather than the average scores over a schedule. The HSE FRI calculator within CrewPlan calculates the average scores over a specified period. Therefore, there needs to be a level of caution when extracting the information.

2.4.1.1: Interpretation of the Fatigue Index (FI)

The following equation was developed to calculate the fatigue index (FI) score (Spencer *et al*., 2006):

2.4.1.2: The Fatigue Index (FI) equation:

$$\mathrm{FI} = 100\,\{1 - (1 - \mathrm{C})\,(1 - \mathrm{J} - \mathrm{T})\}$$

Where by:

C = the cumulative fatigue component;
T = the duty timing component;
J = the job type / breaks component.

The FI value is based on a score between 0-100, which represents the average probability expressed as a percentage of high levels of sleepiness (Spencer *et al*., 2006). The FI score represents the level of fatigue as a percentage that an employee will experience if (s)he completes the allocated weekly roster. A value of 20.7 corresponds to the average from a 4-day working roster (i.e., Day shift, Day shift, Night shift, Night shift, Rest day, Rest day, Rest day, and Rest day) pattern, assuming typical values for the job type / breaks factor (Spencer *et al*., 2006). For example, imagine an employee was scheduled to work a typical weekly roster that generated a FI score of 50. A score of 50 on the FI represents a 50 per cent chance of employees achieving a Karolinska Sleepiness Score (KSS) of 8 or 9 (see, Åkerstedt & Gillberg, 1990, for review). The KSS is a nine-point scale ranging from one (extremely alert) to nine (extremely sleepy, fighting sleep). It has been extensively validated (Harma *et al.*, 2006; Kaida *et al.*, 2006; Lo *et al.*, 2012; Sagaspe *et al.*, 2008) and provides an adequate substitute for direct physiological measurement (Brown, Wieroney, Blair, & Zhu, 2014). In other words, if a score of 45 is obtained for a particular shift then the HSE FRI calculator represents a 45 per cent chance of experiencing high levels of fatigue and that the employee would also struggle by 45 per cent to stay awake during that particular shift (Bowler & Gibson, 2015). However, it is important to highlight that since the FI value is cumulative, this means that for every additional working day the employee completes, their percentage of experiencing higher levels of fatigue would also incrementally increase. According to Stanton *et al.* (2010), a FI score of 45 or below represents good practice within the rail industry (Bell, 2008), as well as working to reduce the risk of fatigue for UK rail operators (ORR, 2012). However, Rangan and van Dongen (2013) argue that what is deemed to be acceptable or unacceptable levels of fatigue using biomathematical model (BMM) output depends purely on where the threshold line is drawn. Therefore, Somvang, Hayward and Cabon (2016) have outlined that a BMM should not be used in isolation to make important decisions, as there are no agreed ’thresholds’ for the HSE FRI calculator. Hursh and van Dongen (2010) have pointed out that advances in BMM’s of fatigue have facilitated systematic investigations of issues in the context of the Fatigue Risk Management System (FRMS). Therefore, caution on both the FI and RI thresholds should be taken as these are but one component of the FRMS for effectively assessing levels of fatigue (ORR, 2016a). According to Sadeghniiat-Haghighi and Yazdi (2015), the FRMS is a comprehensive approach that is based on applying scientific evidence of sleep knowledge to manage workers fatigue. Caldwell, Caldwell, Thompson and Lieberman (2019) state that the FRMS is quickly being adopted by the transportation industry, with the Federal Aviation Administration (FAA) in the US proactively encouraging the airlines to establish an effective science-based fatigue risk management programme. However, the UK's rail industry has already been adopting a science-based FRMS for almost a decade (ORR, 2012; RSSB, 2012).

2.4.1.3: Interpretation of the Risk Index (FI)

The following equation was developed to calculate the risk index (RI) score (Spencer *et al.*, 2006):

2.4.1.4: The Risk Index (RI) equation:

$$\mathrm{RI} = \mathrm{C} \times \mathrm{J} \times \mathrm{T}$$

Whereby:

C = the cumulative fatigue component;
T = the duty timing component;
J = the job type / breaks component.

The risk index (RI) calculator functions by comparing two shifts. A base shift pattern is given a value of 1.0, which is based on the average levels of a risk of an accident occurring or the error attained from previous studies, which examined the working patterns of employees who completed 12 hour shifts on a 2 day, 2 night and 4 rest day schedule within the rail industry (Bowler & Gibson, 2015). In other words, a risk score of 1.0 is based on the average level of risk of accident/error attained in studies on people working 12 hour shifts on that 2 day, 2 night, 4 rest day schedule in the rail sector. Therefore, a risk score of 1.6 would be an increase of 60 per cent of risk on that particular shift schedule. It is important to highlight that in order to effectively use the RI calculator, businesses need to determine the level of risk that their industry considers acceptable based on the type of work being carried out as well as who is actually doing the work, in terms of their mastery (Bowler & Gibson, 2015). For example, if the work is safety critical then the business may decide a risk score beyond a threshold would be unacceptable (Bowler & Gibson, 2015). According to Stanton *et al*. (2010), a RI score of 1.6 or below represents good practice within the rail industry (Bell, 2008), as well as working to reduce the risk of fatigue for UK rail operators (ORR, 2012). However, once again (as previously outlined), the biomathematics models should not be used in isolation to make important decisions, as there are no agreed 'thresholds' for the HSE FRI calculator (Somvang *el al.*, 2016). Therefore, caution on both the FI and RI thresholds should be taken as these are but one component of the FRMS in assessing fatigue (ORR, 2016a).

2.4.2: Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator assessment

The Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator biomathematical model (BMM) allows the end user to enter standardised values such as commuting time, type of job workload, type of job attention, and breaks (see Figure 6). These values serve as the building block parameters for the HSE FRI calculator to be able to generate scores, which are used for determining the calculations of the cumulative effect component, duty timing component, and job type / break component.

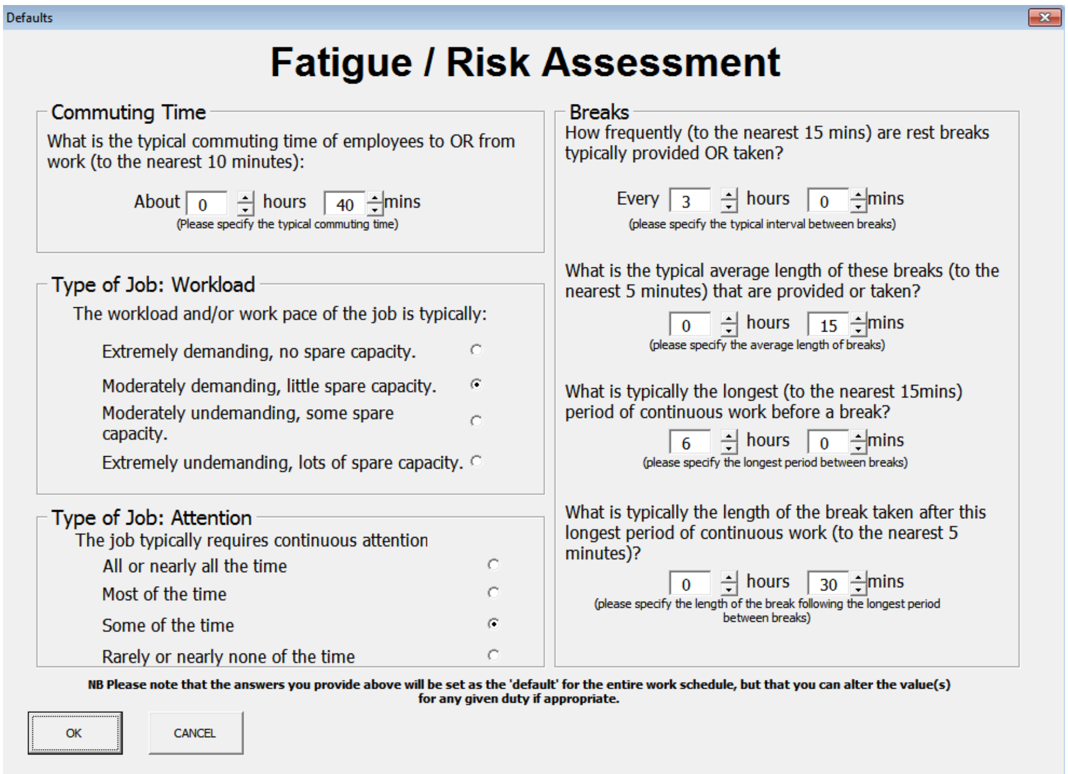


Figure 6: Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator assessment

2.5: Safety incidents in which fatigue could be a contributing factor

Frontline safety critical workers such as train drivers require extensive periods of prolonged concentration and an ability to recall significant quantities of continuously updated and revised information, as well as an ability to fully understand the various traction and route knowledge variations in order to safely operate the unit which transports passengers to their destinations. Section 4.4 outlines some of the information that train drivers must absorb and retain prior to starting their diagram scheduled sheet, e.g., the Signal Passed at Danger (SPAD) notice case board, the weekly operating notice booklet, the late notice case board, and the new notice case board. Therefore, in order for train drivers to be able to obtain a Category B train driving licence under the Train Driving Licences and Certificates Regulations 2010 (TDLCR), a trainee train driver will undergo an extensive two years training programme that includes, but is not limited to, understanding the standards and rules of the mainline railway system, the Train Operating Company's (TOC) unique protocols and procedures, traction knowledge, and route learning/knowledge, etc. As a result, for this thesis/the current study, it was deemed appropriate that the observer had discussions with the occupational safety manager, the head of safety, and the head of resources at Arriva Trains Wales (ATW) to better understand and identify safety incidents categories in which fatigue could be a contributing factor when accounting for the extensive training train drivers receive, as well as ATW's periodical assessment of train driver's competency levels. Based on discussions, it was identified that in the most probable safety incident categories, fatigue could be a contributing factor. As a result, five categories were identified in order of severity; Signal Passed at Danger (SPAD), control systems i.e., Train Protection Warning System (TPWS) activation and Automatic Train Warning System (AWS) slow to cancel, and operational incidents i.e., failed to call and station overrun (see Figure 7, for a visual breakdown). Below is a brief description of the five safety incident categories that were identified.

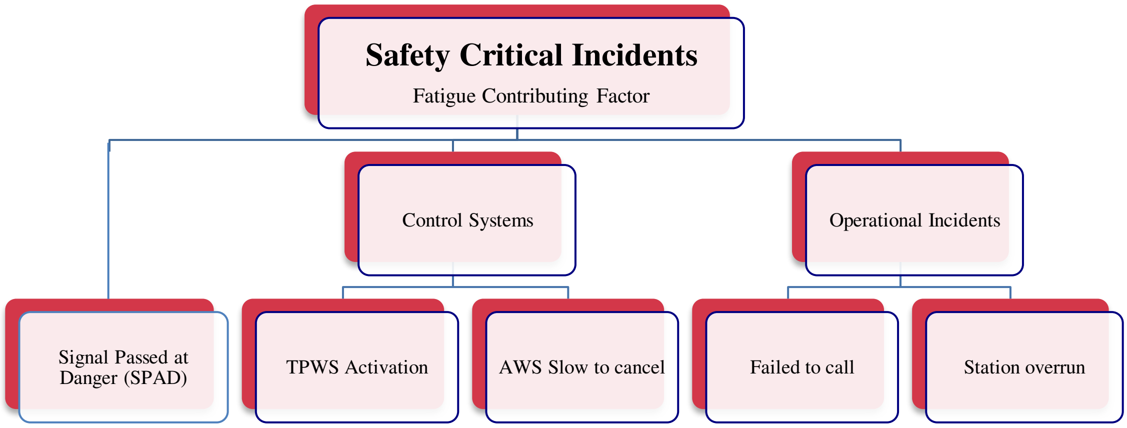


Figure 7: Safety incident group

2.5.1: Signal Passed at Danger (SPAD)

A signal passed at danger – or SPAD – occurs when a train goes past a red signal without the authority of the signaller (RSSB, 2018a). It is important to point out that for the most part, SPADs nowadays have little or no potential to cause harm as they are a result of train drivers misjudging distances whether due to individual traction braking sensitivity setup or low adhesion (e.g., leaves on the line), which usually tends to occur at reduced speed (RSSB, 2016a). In most cases, the trains will stop well within the safety overlap route section, whether by the operating train driver or through the Train Protection Warning System (TPWS) intervention (see Section 2.5.2.1). However, prior to the integration of the TPWS in the early 2000s (RSSB, 2015b), SPADs have been the precursor to some of the most serious fatal train accidents in history, which happened in Purley in 1989, Newton in 1991, Cowden in 1994, Watford in 1996, and at Southall in 1997. The last multi-fatality train accident as the result of a SPAD occurred in October 1999 at Ladbroke Grove (RSSB, 2018b), prior to the nationwide implementation of the TPWS (RSSB, 2015b). However, since the implementation of the TPWS, there have been

4 train collisions and more than 50 derailments caused by SPADs, with the notable near miss incidents that occurred in 2015 at Wootton Bassett (RSSB, 2018b).

2.5.2: Control systems

Extensive discussions with the senior management team at Arriva Train Wales (ATW) revealed that train drivers who may be experiencing lapse in attention due to fatigue are potentially more likely to be involved in control systems safety critical incidents such as Train Protection Warning System (TPWS) activation and Automatic Warning System (AWS) slow to cancel. These two safety critical incidents in which fatigue could be a contributing factor are outlined below.

2.5.2.1: Train Protection Warning System (TPWS) activation

The Train Protection Warning System (TPWS; see Figure 8, for a visual representation) was implemented across the United Kingdom's rail network as an interim measure to reduce the consequences of the safety critical incident - signal passed at danger (SPAD; Moor, 2013). Widespread fitment of the TPWS began in early 2000, in order to meet the requirements of the Railway Safety Regulations 1999 (RSSB, 2015b).

Figure 8: Train Protection Warning System (TPWS)

RSSB (2015b: 8), state that the purpose of TPWS is to stop a train by automatically initiating the brakes when a train has:

- *'passed a signal at danger without authority;*
- *approached a signal at danger too fast;*
- *approached a reduction in permissible speed too fast;*
- *approached buffer stops too fast.'*

It is important to highlight that according to RSSB (2015b), the purpose of the TPWS is not to prevent SPADs, but to reduce the likelihood of a such a safety critical incident from occurring by preventing the unit from reaching a particular point ahead of the signal. There are two types of TPWS, TPWS interventions and TPWS activations. Both of these types have their own unique definition and meaning, which are clearly defined in the rule book. Moor (2013: 8) outlines that these two TPWS' types are defined as;

a) 'TPWS Intervention

A TPWS intervention occurs when the TPWS applies the brakes in the absence of (or prior to) the driver doing so. For example:

- *A train starting against a TPWS-fitted signal at danger without authority will result in an* ***intervention*** *when the train passes the signal;*
- *A driver taking no action to apply the brake on approaching a signal at danger and passing over the overspeed loops too quickly will also result in an* ***intervention****.*

In short, the safety system 'intervenes' if the driver has not taken the appropriate action.

b) TPWS Activation

This occurs when a driver has already applied the brakes before the TPWS operates. For example:

- *A driver might already be braking on the approach to a red signal, but still passes over the overspeed sensor too quickly, resulting in an* ***activation****;*
- *If a train passes a TPWS-fitted signal at danger, despite having applied the brakes in an attempt to stop at it, then an* ***activation*** *results.*

In short, the safety system 'activates' to back up the driver's brake application.

2.5.2.2: Automatic Warning System (AWS) slow to cancel

The Automatic Warning System (AWS) was implemented across the United Kingdom's rail network to provide train drivers with both an audible as well as a visual indicator of whether the distant signal was either clear (i.e., green) or at caution (i.e., either yellow or red; RSSB, 2015b). Without going into the technical specifications of the various AWS equipment components installed on the track (see Figure 9, for a visual representation), under the traction unit (i.e., AWS receiver), and inside the cab (RSSB, 2015b), only a brief description of what the train driver is required to fully understand as well as their responsibilities will be outlined. There are three distinctive AWS indicators inside the cab (see Figure 10) – the AWS audible indicator, the AWS visual indicator, and the AWS/TPWS acknowledgement button. If the train driver fails to acknowledge the AWS/TPWS acknowledgement button e.g., lapse in attention due to fatigue, an emergency brake application will occur and be recorded as an AWS slow to cancel.

Figure 9: Automatic Warning System (AWS) track equipment (permanent magnet and electromagnet)

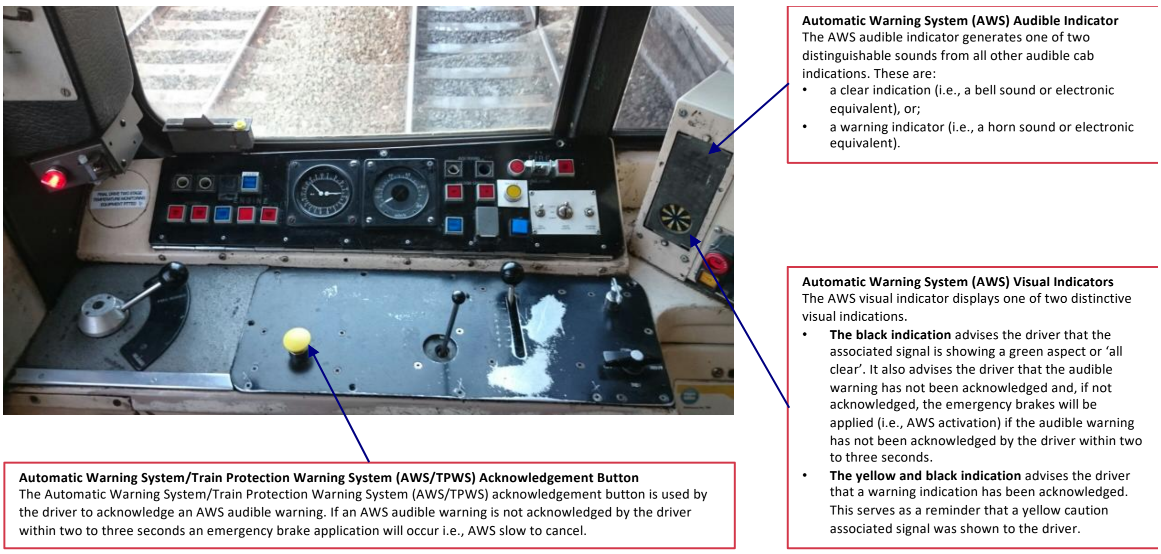


Figure 10: Class 143 Automatic Warning System (AWS) layout, adopted from RSSB (2015b: 6)

2.5.2.2.1 Automatic Warning System (AWS) audible indicator

The Automatic Warning System (AWS) audible indicator generates one of two distinguishable sounds from all other audible cab indications. These are a clear indication (i.e., a bell sound or electronic equivalent), or a warning indicator (i.e., a horn sound or electronic equivalent; RSSB, 2015b).

2.5.2.2.2 Automatic Warning System (AWS) visual indicators

The Automatic Warning System (AWS) visual indicator displays one of two distinctive visual indications (RSSB, 2015b: 6). These are:

- '*The black indication advises the driver that the associated signal is showing a green aspect or 'all clear'. It also advises the driver that the audible warning has not been acknowledged and, if not acknowledged, the emergency brakes will be applied (i.e., AWS activation) if the audible warning has not been acknowledged by the train driver within two to three seconds.*
- *The yellow and black indication advises the driver that a warning indication has been acknowledged. This serves as a reminder that a yellow caution associated signal was shown to the train driver.*'

2.5.2.2.3 Automatic Warning System/Train Protection Warning System (AWS/TPWS) acknowledgement button

The Automatic Warning System/Train Protection Warning System (AWS/TPWS) acknowledgement button is used by the train driver to acknowledge an AWS audible warning (RSSB, 2015b). If an AWS audible warning is not acknowledged by the train driver within two to three seconds an emergency brake application will occur i.e., AWS slow to cancel.

2.5.3: Operational incidents

Extensive discussions with the senior management team at Arriva Train Wales (ATW) revealed that train drivers who may be experiencing lapse in attention due to fatigue are potentially more likely to be involved in operational safety critical incidents such as failed to call, and station overrun. These two safety critical incidents in which fatigue could be a contributing factor are outlined below.

2.5.3.1: Failed to call

According to RSSB (2016b: 5), a failed to call is a 'failure of a train to make a booked station stop in cases where the driver has made no attempt to apply the brake.' At Arriva Trains Wales (ATW), several routes have request only non-booked station stops or booked station stops that are diagram schedule sheet specific e.g., station stop based on the specific station's timetable may only stop hourly despite several trains operating on the same route within said hour. In other words, some stations have far less frequently booked station stops due to them being less busy as a result of being smaller commuter stations. According to RSSB (2010a), failed to call at a station can have a negative impact on safety, passenger journey and company business.

2.5.3.2: Station overrun

RSSB (2016b: 5) state that a station overrun is an 'event in which a train which the driver is attempting to bring to a stand at a station stop proceeds beyond the designated stopping point such that any door(s) intended to be for passenger use at that station is no longer on the operational platform.' According to RSSB (2010a), at least 69 per cent of station overruns are due to error, which includes train drivers, driver managers, unit design and manufacture errors, with the remaining 31 per cent being contributed to low adhesion or a mixture between errors and low adhesion. The Arriva Trains Wales (ATW) franchise operates in routes that are known to have dense foliage (i.e., leaves falling on to the line). Therefore, any lapse in attention or situation awareness due to fatigue could result in station overrun, whether due to low adhesion (e.g., leaves on the line) or non-booked station stops (i.e., request only stops) and booked station stops that are diagram schedule sheet specific (see above section for clarification).

2.6: Chapter summary

The aim of this chapter was to provide an overview of the current Fatigue Risk Management System (FRMS) at Arriva Trains Wales (ATW) as well as to be able to identify and provide an overview of the safety incidents that were deemed to be the most serious in which fatigue could be a contributing factor i.e., Signal Passed at Danger (SPAD), Train Protection & Warning System (TPWS) activation, Automatic Warning System (AWS) slow to cancel, failed to call, and station overrun.

Chapter 3: Secondary analyses of large existing data

3.1: Overview of chapter

The previous chapter discussed the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator for assessing train drivers' level of fatigue. The objective of this chapter is to investigate the effectiveness of the HSE's FRI calculator used at Arriva Trains Wales' (ATW) as part of their Fatigue Risk Management System (FRMS) for monitoring and managing safety incidents in which fatigue could have been a contributing factor. Chapter 3 aims to answer two fundamental questions:

1. The aim of the first study is to investigate whether the present biomathematical model (BMM) for assessing train drivers' level of fatigue at ATW i.e., the HSE's FRI calculator relates to the number of safety incidents in which fatigue could have been a contributing factor.

2. The aim of the second study is to investigate whether restricting train drivers from working on their assigned rest days reduced the number of safety incidents in which fatigue could have been a contributing factor.

3.2: Introduction and rationale

The Fatigue Risk Index (FRI) calculator was commissioned by the Health and Safety Executive (HSE) and developed as an objective biomathematical model (BMM) to represent as simply as possible the build-up of fatigue levels (Rogers, Spencer, & Stone, 1999). The HSE FRI calculator relates to the amount of sleep lost that is likely to be associated with the pattern of work and rest (Spencer, Robertson, & Folkard, 2006). As a result, the HSE FRI calculator was designed to assess the level of fatigue experienced by shift workers based on a set of parameters: cumulative effects; duty timing; and job type / breaks. To ensure the HSE's FRI calculator is utilised correctly, it is vital to ensure that when an individual's shiftwork pattern is examined that both the fatigue index (FI) and risk index (RI) scores are reviewed (Spencer *et al*., 2006). Rangan and van Dongen (2013) argue that what is deemed to be acceptable or unacceptable levels of fatigue or risk using the HSE's FRI calculator output depends purely on where the threshold line is drawn.

Stanton *et al*. (2010) state that a FI score of 45 or below or a RI score of 1.6 or below represents good practice within the rail industry, as well as working to reduce the risk of fatigue for UK rail operators (ORR, 2014). However, Somvang, Hayward and Cabon (2016) have outlined that the HSE's FRI calculator should not be used in isolation to make important decisions, as there are no agreed 'thresholds'. Hursh and van Dongen (2010) have pointed out that advances in biomathematical modelling (BMM) of fatigue has facilitated systematic investigations of issues in the context of the fatigue risk management system (FRMS). Therefore, caution on both the FI and RI thresholds should be taken as these are but one component of the FRMS for effectively assessing levels of fatigue (ORR, 2016a). According to Sadeghniiat-Haghighi and Yazdi (2015), the FRMS is a comprehensive approach that is based on applying scientific evidence of sleep knowledge to manage workers fatigue. ATW have implemented the HSE's FRI calculator into their scheduling system – CrewPlan - with the aim of effectively monitoring and managing fatigue levels in frontline safety critical workers.

3.3: Study 1

The aim of the first study is to investigate whether the present biomathematical model (BMM) for assessing train drivers' level of fatigue at Arriva Trains Wales (ATW) i.e., the HSE's FRI calculator, related to the number of safety incidents in which fatigue could have been a contributing factor i.e., Signal Passed at Danger (SPAD), Train Protection & Warning System (TPWS) activation, Automatic Warning System (AWS) slow to cancel, failed to call, and station overrun.

3.3.1: Methodology

3.3.1.1: Ethical approval

The study received ethics approval from Cardiff University's Ethics Committee (EC.16.06.14.4547). The study conformed to the seventh amendment of the Declaration of Helsinki 1964 (World Medical Association, 2013) and was in accordance with the Data Protection Act 1998 and the General Data Protection Regulation (GDPR).

3.3.1.2: Participants

Safety incident reports of 578 train drivers (550 men and 28 female) ranging in age from 24 – 65 years old (M = 47.13, SD = 7.30) were collected using the Safety Management Information System (SMIS). On average, train drivers had driving experience for 10 years before their safety incident occurred.

3.3.1.3: Statistical analyses

IBM's Statistical Package for the Social Sciences (SPSS) version 23 for Mac was used to analyse the data. Three statistical procedures were carried out on the data; descriptive analyses, cross tabulation, and chi-square test of independence to examine the relationships between independent variables.

3.3.1.4: Materials and procedure

All safety incident groups in which fatigue could have been a contributing factor were extracted from SMIS. The safety incident types identified were; SPAD, TPWS activation, AWS slow to cancel, failed to call, and station overrun. Using ATW's CrewPlan scheduling system, train drivers fatigue index (FI) and risk index (RI) scores, as well as other relevant roster information, was extracted. CrewPlan went live 1st June 2010. As a result, this study only examined safety incident reports between 1st June 2010 – 31st December 2016.

3.3.2: Results

A total of 901 safety incidents were recorded between 1st June 2010 – 31st December 2016 within the Safety Management Information System (SMIS). From the 901 recorded safety incidents, only the fatigue index (FI) scores and risk index (RI) scores of 64.2 per cent (n = 578) of train drivers were accessible for analyses from Arriva's scheduling system CrewPlan (see Figure 11). The other 35.8 per cent of recorded safety incidents did not contain sufficient information to be able to identify the train driver. As can be seen from Figure 11, 35.6 per cent (n = 206) of all recorded safety incidents were classified as Train Protection & Warning System (TPWS) activation, followed by; Automatic Warning System (AWS) slow to cancel with 24.7 per cent (n = 143), station overrun with 20.6 per cent (n = 119), Signal Passed at Danger (SPAD) with 9.7 per cent (n = 56), and the least recoded safety incident classification was failed to call with 9.3 per cent (n = 54).

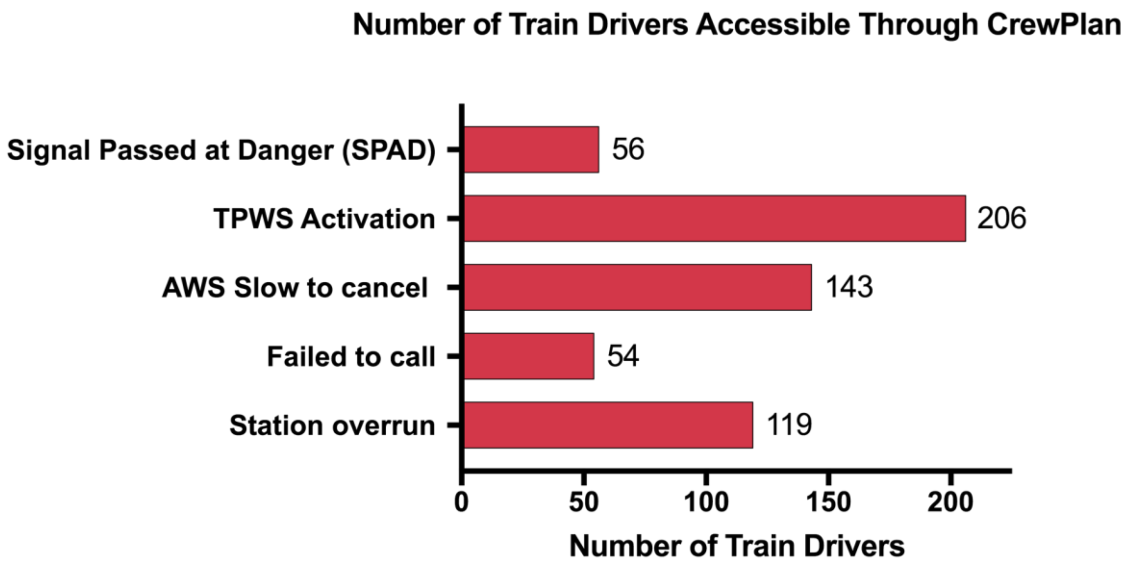


Figure 11: Train drivers CrewPlan accessibility

The number of recorded safety incidents of train drivers involved in safety incidents showed an increase of 150.9 per cent when comparing the total numbers from the year 2011 (n = 57) to the year 2016 (n = 143) across all five groups in which fatigue could have been a contributing factor i.e., SPAD, TPWS activation, AWS slow to cancel, failed to call, and station overrun (see Figure 12). A breakdown of each safety incident group can be seen in

Figure 13. It is important to note that 5.4 per cent (n = 31) of all accessible train driver's safety incidents which corresponded to the period 1st June 2010 - 31st December 2010 were excluded from Figure 12 and Figure 13 as this time period significantly underrepresents the total number of safety incidents in which fatigue could have been a contributing factor. This was due to the fact that CrewPlan within ATW did not officially go live until the 1st June 2010 and therefore was not able to extract train driver's fatigue index (FI) and risk index (RI) scores.

Figure 12: Total number of safety incidents across all five groups in which fatigue could have been a contributing factor

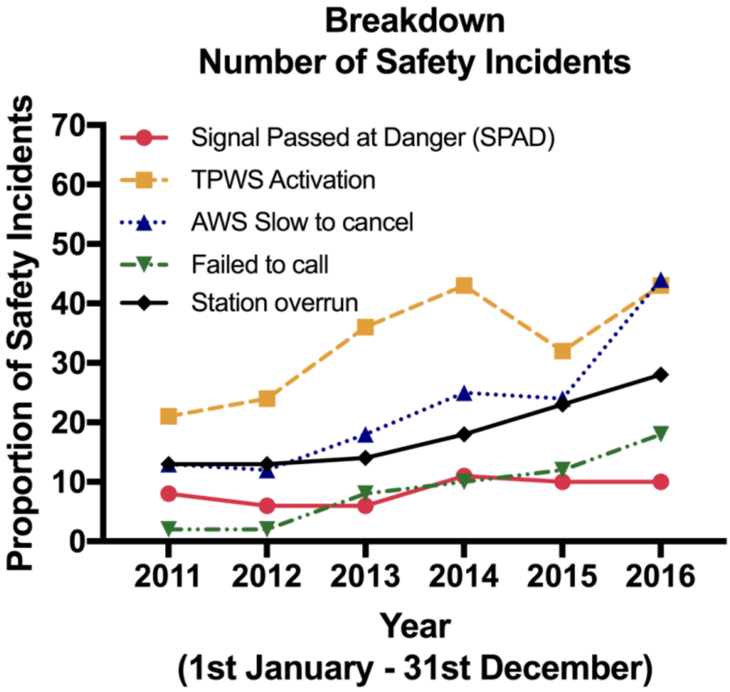


Figure 13: A breakdown of the total number of safety incidents of all five groups in which fatigue could have been a contributing factor.

3.3.2.1: Comparing CrewPlan to the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator

Replicability and validity tests were carried out to compare the fatigue index (FI) and risk index (RI) scores that were generated by ATW's CrewPlan scheduling system and the Fatigue Risk Index (FRI) calculator (i.e., biomathematical model) developed by Spencer *et al*. (2006) for the Health and Safety Executive (HSE), using the shiftwork roster of three randomly selected train drivers. According to K. Robertson (personal communication, February 24, 2017), a principal scientist at QinetiQ was able confirm that the CrewPlan scheduling system at ATW uses the exact same formulas and calculations from the HSE's FRI calculator. To demonstrate that the HSE FRI calculator for the indices are indeed proprietary information of the Health and Safety Executive (HSE), the below communication from K. Robertson (personal communication, February 24, 2017) outlined that:

'The module that QinetiQ provides to them is a replica of the Excel spreadsheet that is found on the HSE website [...]. As the formulas used in the calculation of the Indices are proprietary information I'm afraid I can't give you any more details about the calculations.'

However, it was found that only the RI calculations generated the same scores, and that the FI scores within ATW's CrewPlan and HSE FRI calculator generated slightly different scores, despite all input parameters, dates, and time of duties being identical. A possible explanation was highlighted by P. Ayres (personal communication, February 22, 2017), who stated that:

'Differences between fatigue and risk scores produced from Crewplan and those produced from the standalone spreadsheet are not unusual. It is usually because the standalone spreadsheet evaluates a 'snapshot' of shift data pasted in, while the Crewplan version provides cumulative scores from several weeks of roster data prior to the dates requested. This can normally be observed through the standalone spreadsheet starting with low fatigue and risk scores – it always starts from 0, as if from a prolonged period of rest. The Crewplan version normally starts with much higher score because of cumulative fatigue and risk from the roster data prior to the dates being analysed.

Unfortunately, Crewplan doesn't actually calculate fatigue and risk scores itself [...]. Crewplan supplies shift times to a self-contained fatigue and risk module, which is licensed from QinetiQ. It is the QinetiQ module that outputs the fatigue and risk scores back to Crewplan.'

3.3.2.2: Arriva Train Wales' (ATW) Health and Safety Executive's (HSE) Fatigue Risk Index (FRI)

Each depot of the Arriva Trains Wales (ATW) franchise has input parameters that are exclusively based on average depot predictions. In other words, these parameters are based on train drivers as a cohort, rather than train driver's individual characteristics, individual diagram route workload, shift intensity, rolling stock condition, individual cabin ergonomics, etc. Therefore, each depot at ATW implements their own acceptable configuration and calibration parameters into CrewPlan, based on individual operational needs. However, the scheduling team has oversight to ensure that all frontline safety critical workers such as train drivers adhered to Arriva Trian Wales' (2017) fatigue recommended management thresholds of FI = 45 and RI = 1.6 as part of ATW's Fatigue Risk Management System (FRMS). Table 2 outlines each depot's FRMS configuration and calibration parameters (see Table 2).

Fatigue parameter calibration from ATW's scheduling system, CrewPlan, which integrates the HSE's FRI calculator revealed that all depots across the franchise defined their fatigue index (FI) and risk index (RI) upper limit thresholds in accordance with ATW's (2017) Fatigue Risk Management System (FRMS) recommendation i.e., FI = 45, and RI = 1.6. In addition, it was also found that CrewPlan granted authorisation to the resource management team to be able to validate the FI and RI scores when rotating a diagram turns. Furthermore, it was also found that CrewPlan's configurations indicated that all diagram workload in frontline safety critical workers were classified as 'extremely demanding, no spare capacity' as well as that all types of jobs were classified as 'job typically requires continuous attention: All or nearly all the time', with both classifications presenting consistencies across the franchise. However, inconstancies were found in ATW's integration and configuration of the HSE's FRI calculator.

A closer examination of the fatigue configuration parameters within ATW's CrewPlan revealed that Cardiff depot (CFD) and Treherbert depot (TTD) did not 'treat timed stand-by-spares as working days'. This is problematic when considering that in frontline safety critical worker, such as train drivers both 'spare' or 'stand-by' diagram turns still require train drivers to report for duty and potentially catch a turn(s) i.e., assigned a diagram or partial diagram. Therefore, these depots will fail to correctly calculate and generate frontline safety critical worker, such as train drivers' fatigue index (FI) and risk index (RI) scores. In addition, inconsistencies were also found in the parameters set up for the 'typical commuting time of employee to/from work'. For example, Machynlleth depot (MND) has 45 minutes entered into the HSE's FRI calculator within ATW's CrewPlan for commuting time, while Holyhead depot (HDD) has 15 minutes. In contrast, Cardiff depot (CFD) has 35 minutes while Pwllheli depot (PWD) has 45 minutes. Ideally, commuting time should be unique to the frontline safety critical worker as well as incorporating population density of the region, traffic congestion, and parking facilities.

3.3.2.3: Frequency of safety incidents against time of safety incidents

At present, Arriva Trains Wales (ATW) does not currently have policies in place that categorises what are considered to be the various shift types. However, according to S. Handley (personal communication, June 26, 2018) (see Figure 14)

> *'rule of thumb I have always used with the unions is as follows: High risk shifts starting between 00:01 – 04:59 and any shifts working through the night (like shed turns), AM turns booking on 05:00 – 10:00, Mid shifts booking on 11:00 - 13:00, which finish before 20:00. PM shifts booking on after 14:00 – 20:00. Anything after 20:00 classes as a night shift.'*

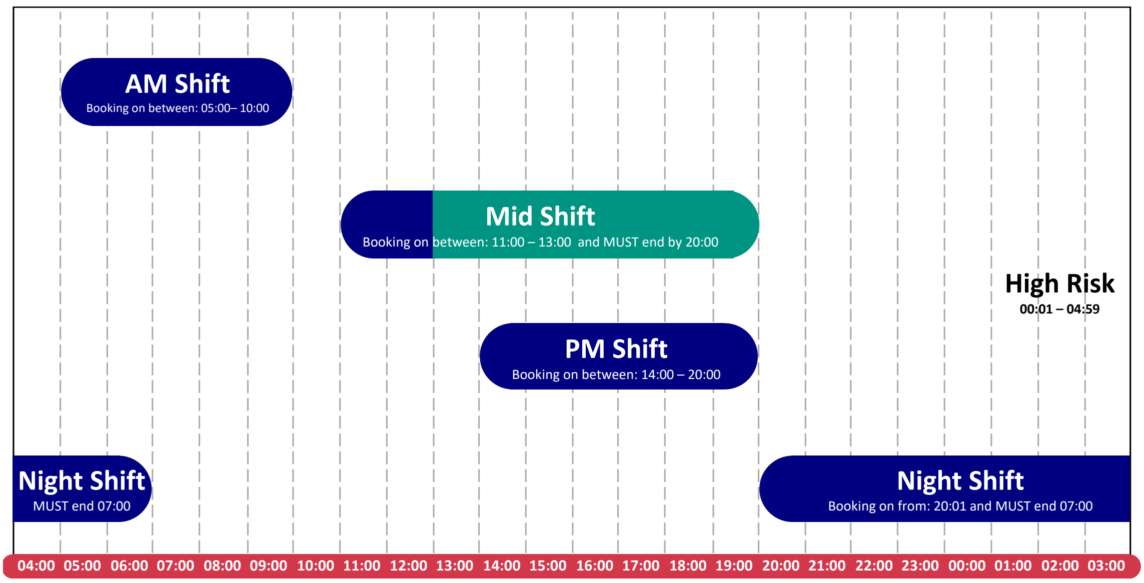


Figure 14: Shift types

	North Depot							South Depot				
	CED	CHD	HDD	LJD	MND	PWD	SQD	CFD	CND	CVD	RHD	TTD
Fatigue – Risk Index Threshold												
Fatigue Should Not Exceed	45	45	45	45	45	45	45	45	45	45	45	45
Fatigue Must Not Exceed	45	45	45	45	45	45	45	45	45	45	45	45
Risk Should Not Exceed	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6
Risk Must Not Exceed	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6
Depot Parameters												
Specify parameters by individual link												
Treat timed stand-by-spares as working days	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	
Validate Fatigue/Risk index scores when rotating a roster	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Type of Job Workload												
Workload and/or work pace of the job typically:												
Extremely demanding, no spare capacity	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Moderately demanding, little spare capacity												
Moderately undemanding, some spare capacity												
Extremely undemanding, lots of spare capacity												
Type of Job: Attention												
Job typically requires continuous attention:												
All or nearly all the time	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Most of the time												
Some of the time												
Rarely or nearly none of the time												
Commuting Time:												
Typical commuting time of employee to/from work:	00:45	00:45	00:15	00:45	00:45	00:45	00:45	00:35	00:35	00:35	00:30	00:30
Breaks												
How frequently are rest breaks typically provided or taken?	03:00	03:00	03:00	03:00	03:00	03:00	03:00	03:10	03:10	03:00	03:10	03:10
Typical average length of breaks	00:30	00:30	00:30	00:30	00:20	00:20	00:30	00:30	00:20	00:20	00:20	00:20
Typically the longest period of continuous work before break	04:15	04:15	04:15	04:15	04:15	04:15	04:15	04:15	04:15	04:30	04:30	04:30
Typical length of the break taken after this longest period of continuous work	00:30	00:30	00:30	00:30	00:30	00:30	00:30	00:40	00:40	00:40	00:40	00:40

Table 2: Arriva Trains Wales' (ATW) Fatigue Risk Management System (FRMS) individual depot parameters within CrewPlan

Note:

North depots: Crewe (CED); Chester (CHD); Holyhead (HDD); Llandudno Junction (LJD); Machynlleth (MND); Pwllheli (PWD); and Shrewsbury (SQD).

South depots: Cardiff (CFD); Carmarthen (CND); Cardiff Valleys (CVD); Rhymney (RHD); and Treherbert (TTD).

According to The Working Time Regulations 1998, the night period is between 11pm to 6am, unless a contract between the worker and employer agree to a different night period. If an alternative contract is agreed in writing, it must be 7 hours long and include midnight to 5am. ATW unofficially recognises night shift starting from 20:00 and any shifts working through the night, such as turn 867 in the shed that starts at 23:00 and finishes at 07:00. Figure 14 also represents this turn, which is the latest at ATW.

Based on observed frequencies of the time of day the safety incidents occurred, it was found that 61.7 per cent (n = 418) of all safety incidents took place between 08:00 – 20:00 hours. These findings are not that surprising as there are more operational rolling stock fleet (i.e., trains) during 'normal' working hours (09:00 – 17:00). When looking at a 2-hour categorical breakdown of the time of day frequencies the safety incidents occurred (see Figure 15), there are no obvious observed spikes, which seems to indicate that perhaps there may not be a specific time of day when safety incidents are more prominent. No further analyses were carried out to investigate whether safety incidents were more prominent at a shift type since ATW does not have a pre-defined policy in place. Therefore, splitting the data would not be an accurate presentation based on a 'rule of thumb' agreement between ATW and the unions i.e., the Associated Society of Locomotive Steam Enginemen and Firemen (ASLEF) and the National Union of Rail, Maritime and Transport (RMT).

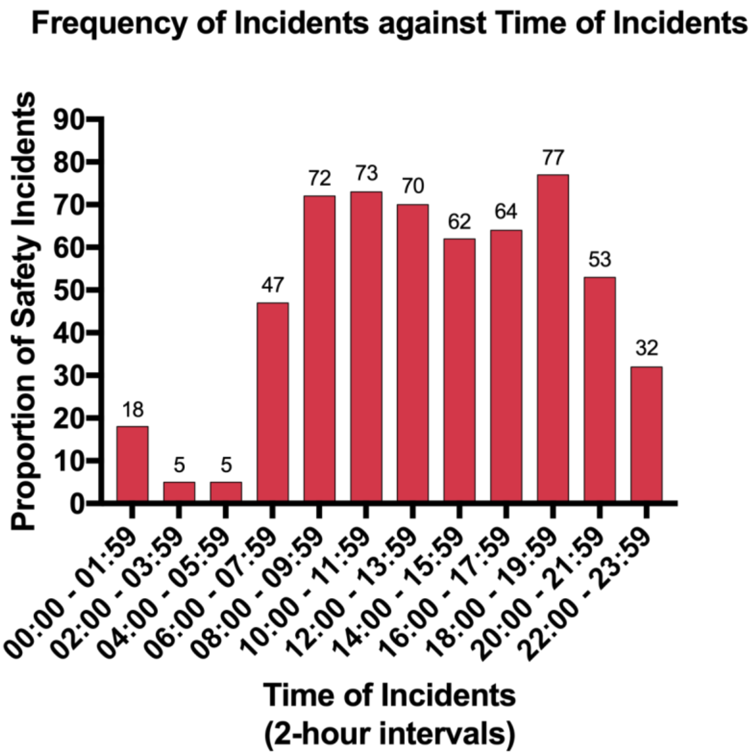


Figure 15: Frequency of safety incidents against time of safety incidents

3.3.2.4: Number of safety incidents against time of incidents

Further analyses were carried out to investigate the frequency of safety incidents against time of incidents, as 61.7 per cent (n = 418) of all safety incidents took place between 08:00 – 20:00 hours. When looking at a 2-hour categorical breakdown of the frequency of safety incidents against the time of incidents (see Figure 16), there are was an obvious observed spike for TPWS Activation. However, a chi-square goodness of fit test, χ^2 (5, n = 162) = 6.22, p = .285 indicated no statistical differences for the TPWS Activation safety incidents against the time of incidents. No further analyses were carried out to explore frequency of safety incidents against time of incidents.

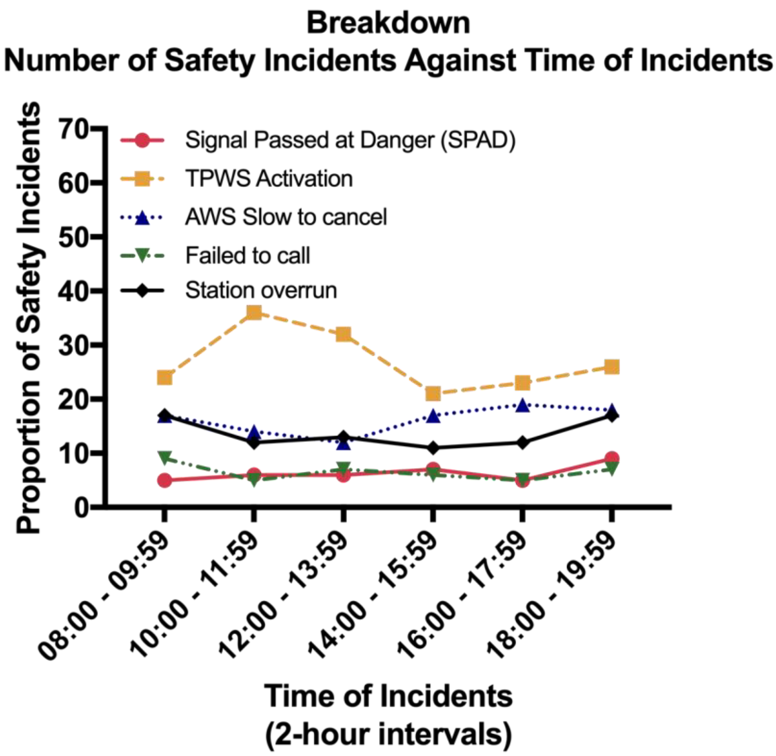


Figure 16: Breakdown of the frequency of safety incidents against the time of incidents

3.3.2.5: Number of safety incidents against length of train driving experience

Further analyses were carried out to investigate the frequency of safety incidents against train driving experience before their safety incident occurred. Train driving experience in this instance is defined by the number of years a train driver has had since obtained their train driver's licence and certificate, which is issued by the Office of Rail and Road (ORR) in accordance to the Train Driving Licences and Certificates Regulations 2010 (TDLCR). When train driving experiences were categorised into 5-unit categories, a chi-square goodness of fit test, χ^2 (5, n = 578) = 517.18, p < .001 indicated statistical differences in the number of safety incidents against length of train driving experience (see Figure 17).

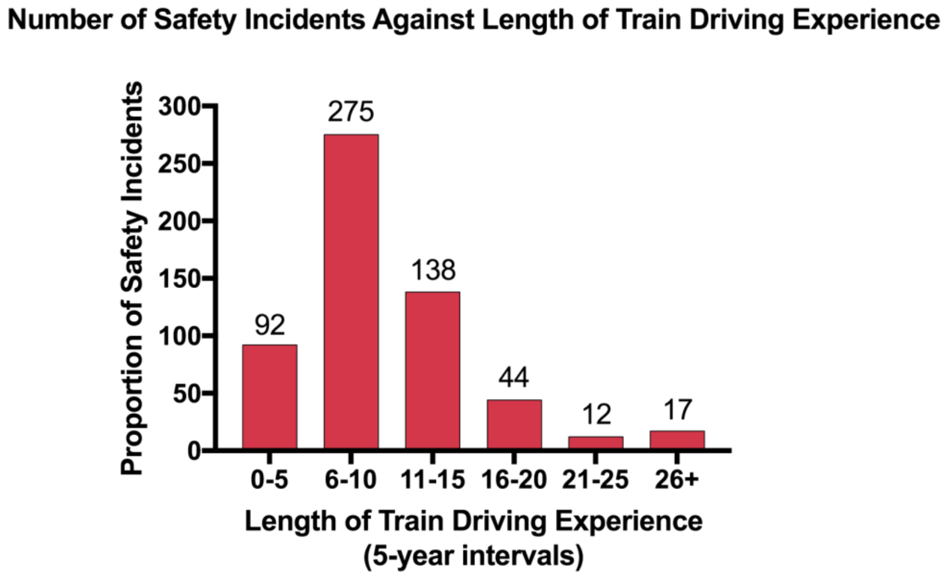


Figure 17: Number of safety incidents against length of train driving experience (5-unit intervals)

In these analyses it was found that 47.6 per cent (n = 275) of train drivers involved in safety incidents had between 6 – 10 years of train driving experience, followed by 23.9 per cent (n = 138) of train drivers who had between 11 – 15 years of train driving experience, followed by 15.9 per cent (n = 92) of train drivers who had between less than a year – 5 years of train driving experience, followed by 7.6 per cent (n = 44) of train drivers who had between 16 – 20 years of train driving experience, followed by 2.9 per cent (n = 17) of train drivers who had more than 26 years of train driving experience, and finally followed by 2.1 per cent (n = 12) of train drivers who had between 21 – 26 years of train driving experience.

3.3.2.6: Breakdown of number of safety incidents against length of train driving experience

Further analyses were carried out to investigate the types of safety incidents in which fatigue could have been a contributing factor against length of train driving experience. Looking at the train drivers who were involved in safety incidents that had between 6 – 10 years of train driving experience, a chi-square goodness of fit test, χ^2 (4, n = 275) = 90.26, p < .001 indicated statistical differences in the type of safety incidents for train drivers who had between 6 – 10 years of train driving experience (see Figure 18).

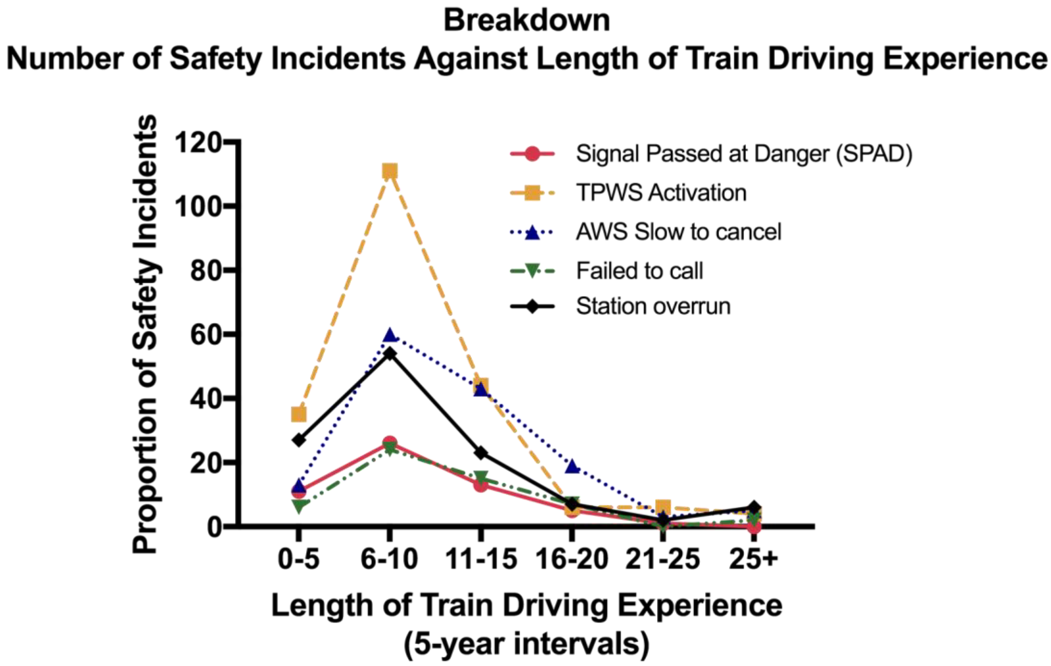


Figure 18: Breakdown of number of safety incidents against length of train driving experience (5-unit intervals)

In these analyses it was found that 40.4 per cent (n = 111) of train drivers who had between 6 – 10 years of train driving experience were involved in TPWS Activation safety incidents, followed by 21.8 per cent (n = 60) of train drivers involved in AWS Slow to cancel safety incidents, followed by 19.6 per cent (n = 54) of train drivers involved in station overrun safety incidents, followed by 9.5 per cent (n = 26) of train drivers involved in Signal Passed at Danger (SPAD) safety incidents, and 8.7 per cent (n = 24) of train drivers involved in failed to call safety incidents. Based on these findings, perhaps the consideration of implementing tailored refresher TPWS Activation after 5 years of train driving experience might be of significant benefit in reducing the number of safety incidents within a revised Safety, Training and Update Day (STUD) programme at ATW.

3.3.2.7: Fatigue

Analysis from all 578 safety incidents have identified that 99.8 per cent (n = 577) of train drivers' fatigue index (FI) scores fell within or below the identified guideline threshold of 45 by Stanton *et al.* (2010) to represent good practice within the rail industry, as well as working to reduce the risk of fatigue for UK rail operators (ORR, 2012) (see Figure 19). A chi-square goodness of fit test χ^2 (1, n = 578) = 1.00, $p < .001$, indicated a statistical difference, which was not surprising since only 0.2 per cent (n = 1) of FI scores fell above the identified guideline threshold of 45. It is important to note that in any given working week, approximately 1 per cent of scheduled train drivers will exceed the recommended working week fatigue index score threshold of 45 (see Stanton *et al*., 2010, for review).

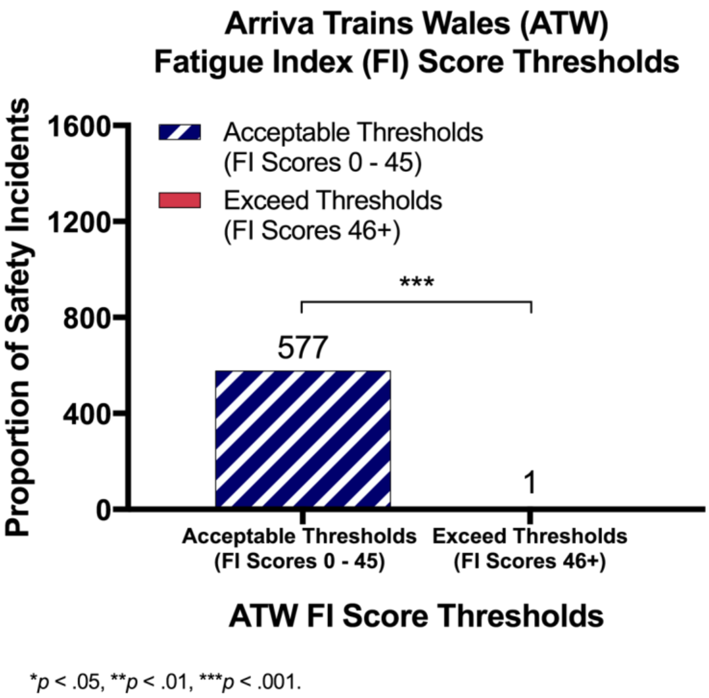


Figure 19: Arriva Trains Wales (ATW) fatigue index (FI) score thresholds

3.3.2.7.1 Categorising Fatigue Index (FI) Scores (5-units)

When the threshold range (n = 578) was further broken down into smaller proportion of fatigue index (FI) scores that fell into each 5-unit category from 0 – 50, it can be seen from Figure 20 that a great proportion of safety incidents occurred when train drivers' FI scores were at the lower end of the rail industry's good practice recommendations. To illustrate this further, the first four 5-unit categories; 0 – 5 (n = 161), 5 – 10 (n = 207), 10 – 15 (n = 108), and 15 – 20 (n = 63), respectively represented 92.8 per cent (n = 539) of all safety incidents between 1st June 2010 – 31st December 2016, which are significantly less than ATW's FI score thresholds of 45, which at that moment in time was also within the rail industry's good practice recommendations (Stanton *et al.*, 2010). However, it is important to state that according to Somvang *et al.* (2016), there are no agreed 'thresholds' for the HSE's FRI calculator. Therefore, caution on the RI thresholds should be taken, as these are but one component of the Fatigue Risk Management System (FRMS) in assessing fatigue (ORR, 2016a).

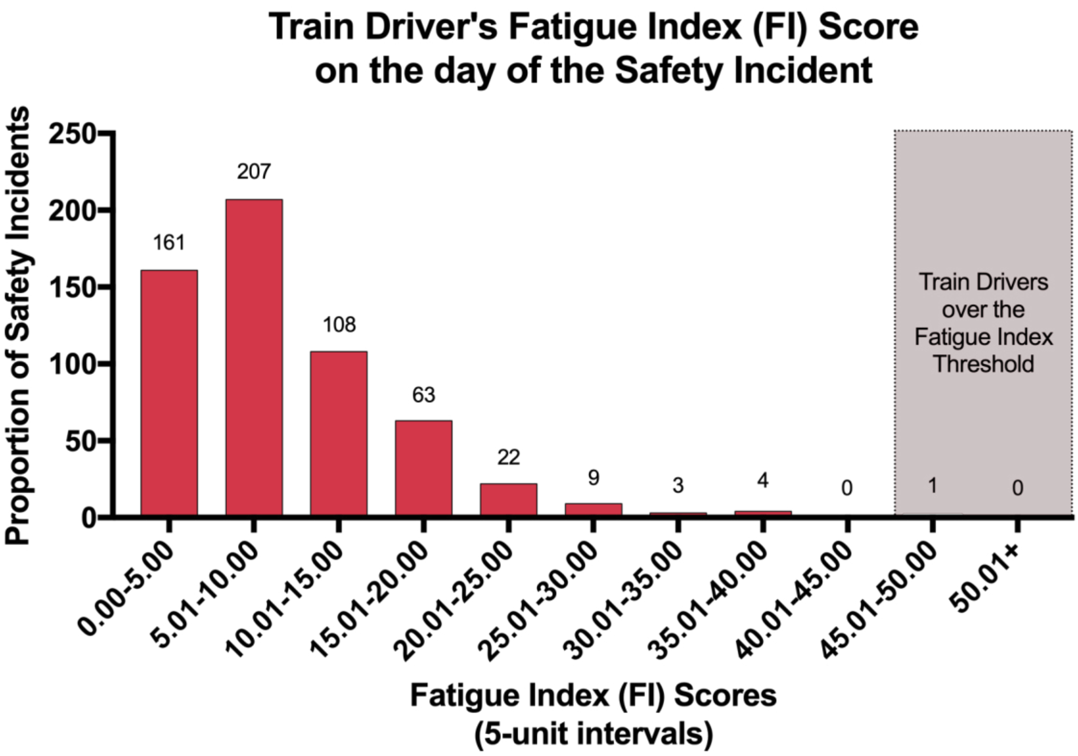


Figure 20: Train driver's fatigue index (FI) scores (5-unit intervals)

3.3.2.8: Risk

Analysis from all 578 safety incidents have identified that 97.9 per cent (n = 566) of train drivers' risk index scores fell below the 1.6 threshold range (see Figure 21). These thresholds have been identified to represent good practice within the rail industry (Stanton *et al.*, 2010) for planned rosters (Bell, 2008), as well as working to reduce the risk of fatigue for UK rail operators (ORR, 2012). A chi-square goodness of fit test χ^2 (1, n = 578) = 1.00, p < .001, indicated a statistical difference, which was not surprising since only 2.1 per cent (n = 12) of RI scores fell above the identified guideline threshold of 1.6.

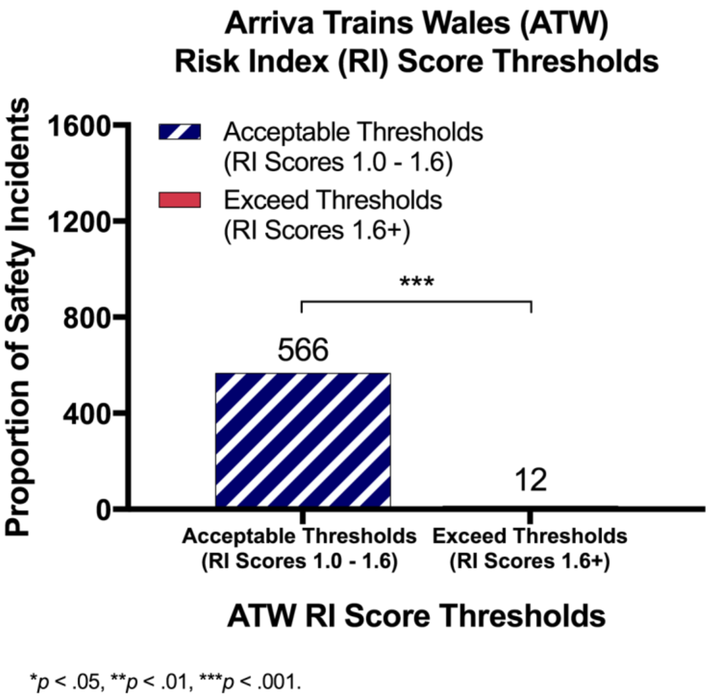


Figure 21: Arriva Trains Wales (ATW) risk index (RI) score thresholds

3.3.2.8.1 Categorising Risk Index (RI) Scores (0.2-unit intervals)

When the threshold range (n = 578) was further broken down into smaller proportion of risk scores that fell into each 0.2-unit category from 0 – 3, it can be seen from Figure 22 that a great proportion of safety incidents occurred when train drivers' risk index (RI) scores were between 0.6 – 1.2, which is within ATW's RI scores as well as the rail industry's good practice recommendations of 1.6 RI score threshold (Bell, 2008; ORR, 2012; Stanton *et al.*, 2010). However, once again, it is important to state that according to Somvang *et al*. (2016), there are no agreed 'thresholds' for the FRI calculator. Therefore, caution on the RI thresholds should be taken as these are but one component of the Fatigue Risk Management System (FRMS) in assessing fatigue (ORR, 2016a).

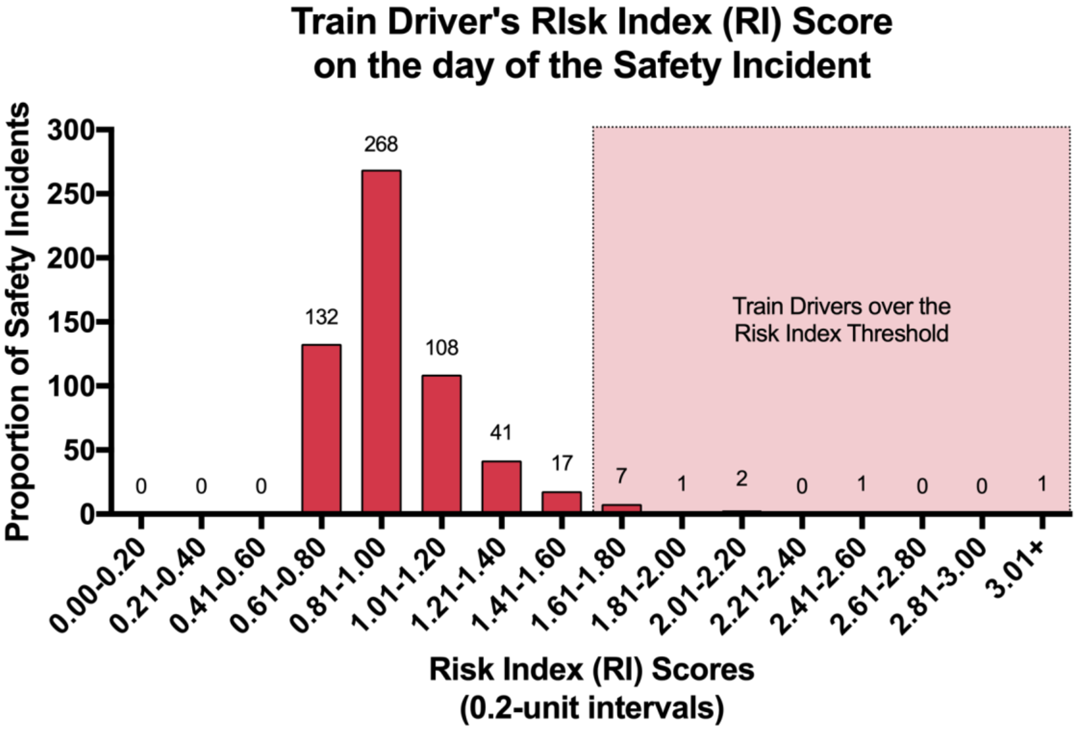


Figure 22: Train driver's risk index (RI) scores (0.2-unit intervals)

3.3.3: Discussion

The aim of this study was to investigate whether the present biomathematical model (BMM) for assessing train drivers' level of fatigue at Arriva Trains Wales (ATW) i.e., the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator related to the number of safety incidents in which fatigue could have been a contributing factor i.e., Signal Passed at Danger (SPAD), Train Protection & Warning System (TPWS) activation, Automatic Warning System (AWS) slow to cancel, failed to call, and station overrun. Analysis from the 578 train drivers who were involved in safety incidents in which fatigue could have been a contributing factor revealed that 99.8 per cent (n = 577) of Fatigue Index (FI) scores fell within or below Arriva Train Wales (ATW) threshold of 45 as well as 97.9 per cent (n = 566) of Risk Index (RI) scores falling below ATW's threshold range of 1.6. However, Bowler and Gibson (2015) identified that the HSE's FRI calculator should be used with caution, as BMMs for both the FI and RI calculations are based on group data, with Somvang *et al*. (2016) stating that there are no agreed 'thresholds' for the HSE's FRI calculator. Therefore, restraint on the FI thresholds should be taken as these are but one component of the Fatigue Risk Management System (FRMS) in assessing fatigue and risk (ORR, 2016a). As a result, there scores were presumed by ATW to represent good practice within the rail industry (Bell, 2008; ORR, 2012; Stanton *et al*., 2010). These findings seem to indicate that train driver's FI and RI scores do not relate to safety incidents in which fatigue could have been a contributing factor, as safety incidents occurred despite 99.8 per cent of FI and 97.9 per cent RI scores were below ATW's threshold range.

One of the biggest limitations of the HSE's FRI calculator at ATW is that input parameters are exclusively based on average depot predictions. In other words, these parameters are based on train drivers as a cohort, rather than train driver's individual characteristics, individual diagram route workload, shift intensity, rolling stock condition, individual cabin ergonomics, etc., as each train driver is unique, and thus will have different fatigue thresholds, as well as different abilities to withstand high levels of fatigue (Bowler & Gibson, 2015). Somvang *et al*. (2016) state that BMM primary application is to assess or compare shifts rather than the characteristics or likely response of an individual. Therefore, the HSE's FRI calculator is not able to take into consideration individual factors (e.g., lifestyle, age, diet, illness, mental conditions, time difference between waking up and clocking in, etc.) or environmental factors (e.g., family circumstance, domestic responsibilities, sleep environment, weather conditions; RSSB, 2012). Moreover, the HSE's FRI calculator generates cumulative fatigue only from the period the driver clocks into work and out (i.e., On Duty, and Off Duty) (Spencer *et al*., 2006).

In this study it was also found that 61.7 per cent (n = 418) of all safety incidents took place between 08:00 – 20:00 hours. A closer examination revealed that TPWS Activations safety incidents were more prominent. However, no statistical significance was found when solely exploring TPWS Activation against time of day incidents In this study it was also found that the largest proportion of safety incidents in which fatigue could have been a contributing factor, 47.6 per cent (n = 275) occurred when train drivers had between 6 – 10 years of train driving experience. Clancy (2007) has outlined that for safety critical operational roles, such as train drivers – there is a danger that rail companies either ‘over-train’ on skills that are of low risk or importance, or ‘under-train’ in areas that are of critical importance. However, Bonsall-Clarke (2012) has identified that an examination of operational incidents and accidents have shown that poor use of non-technical skills (NTS) i.e., situational awareness, conscientiousness, communication, decision making and action, co-operation and working with others, workload management, and self-management (see Bonsall-Clarke, 2012, for review) are contributing factors towards incidents and accidents occurring. Based on these findings, perhaps refining and implementing both revised tailored NTS and areas of critical importance training courses after 5 years of train driving might be potential beneficial to further reduce the number of safety incidents. Further secondary analyses are needed to investigate whether restricting train drivers from working on their assigned rest days based on a naturally occurring intervention reduced the number of safety incidents in which fatigue could have been a contributing factor.

3.3.4: Study summary

The aim of this study was to investigate whether the present biomathematical model (BMM) for assessing train drivers’ level of fatigue at Arriva Trains Wales (ATW) i.e., the Health and Safety Executive’s (HSE) Fatigue Risk Index (FRI) calculator related to the number of safety incidents in which fatigue could have been a contributing factor In this study it was found that the current Health and Safety Executive’s (HSE) Fatigue Risk Index (FRI) calculator for assessing train drivers’ level of fatigue used at Arriva Trains Wales (ATW) may not be an effective or accurate predictor of train driver’s fatigue or risk levels, since analyses of all 578 accessible safety incidents in which fatigue could have been a contributing factor only identified 0.2 per cent (n = 1) and 2.1 per cent (n = 12) of FI scores and RI scores, respectively, that exceeded ATW's Fatigue Risk Management System (FRMS) recommended thresholds of FI score = 45 and RI score = 1.6 (ATW, 2017a). As a result, ATW may need to carefully consider and evaluate with the appropriate urgency whether their current biomathematical model (BMM) i.e., the HSE's FRI calculator is effective at predicting safety incidents in which fatigue could have been a contributing factor in frontline safety critical workers. In addition, it was found that TPWS Activation was the most prominent safety incident for train drivers who had between 6 – 10 years of driving experience. Therefore, it is recommended that going forward, Transport for Wales Rail Services' (TfWRS) Safety, Training and Update Day (STUD) programme reflects and addresses these high TPWS Activation safety incidents. Further analyses are needed to investigate whether restricting train drivers from working on their assigned rest days based on a naturally occurring intervention reduced the number of safety incidents in which fatigue could have been a contributing factor.

3.4: Study 2

The aim of the present study is to investigate whether restricting train drivers from working on their assigned rest days based on a naturally occurring intervention reduced the number of safety incidents in which fatigue could have been a contributing factor.

3.4.1: Methodology

3.4.1.1: Ethical approval

The study received ethics approval from Cardiff University's Ethics Committee (EC.16.06.14.4547). The study conformed to the seventh amendment of the Declaration of Helsinki 1964 (World Medical Association, 2013) and was in accordance with the Data Protection Act 1998 and the General Data Protection Regulation (GDPR).

3.4.1.2: Participants

Safety incident reports of 278 train drivers (266 men and 12 female) ranging in age from 24 – 63 years old (M = 47.46, SD = 7.48) were collected using the Safety Management Information System (SMIS). On average, train drivers had driving experience for 10 years before their safety incident occurred.

3.4.1.3: Statistical analyses

IBM's Statistical Package for the Social Sciences (SPSS) version 23 for Mac was used to analyse the data. three statistical procedures were carried out on the data; descriptive analyses, cross tabulation and chi-square test of independence.

3.4.1.4: Materials and procedure

All safety incident groups in which fatigue could have been a contributing factor were extracted from SMIS. The safety incident types identified were; SPAD, TPWS activation, AWS slow to cancel, failed to call, and station overrun. Using ATW's CrewPlan scheduling system, train drivers fatigue index (FI) and risk index (RI) scores, as well as other relevant roster information, was extracted. CrewPlan went live 1st June 2010. As a result, this study only examined safety incident reports between 1st June 2010 – 31st December 2016.

3.4.1.5: Procedure

Between Monday 14th March 2016 – Monday 19th September 2016, ATW restricted train drivers from working their assigned rest days i.e., train drivers were not permitted to work additional shifts on their allocated days off. This meant that during this time period, all train drivers had a minimum of two mandatory rest days per week. Therefore, the observed frequency of safety incidents during this naturally occurring intervention time period i.e., 14th March 2016 – 19th September 2016 was compared across the same dates in previous years going as far back as 2011.

3.4.2: Results

The aim of this study was to investigate whether restricting train drivers from working on their assigned rest days based on a naturally occurring intervention reduced the number of safety incidents in which fatigue could have been a contributing factor. A chi-square goodness of fit test, χ^2 (4, n = 278) = .667, p > .05, indicated no statistical differences in safety incident frequencies throughout the six time periods (see Figure 23 and Figure 24). However, despite there being no statistical significance between all six time periods, during the restriction period, the observed frequency of safety incidents had increased by 47 per cent (n = 83) when comparing to the same dates in the previous year (n = 44). A chi-square goodness of fit test, χ^2 (1, n = 127) = 1.00, p < .001, indicated statistical differences in safety incident frequencies between 14th March – 19th September 2016 and 14th March – 19th September 2015 (see Figure 25).

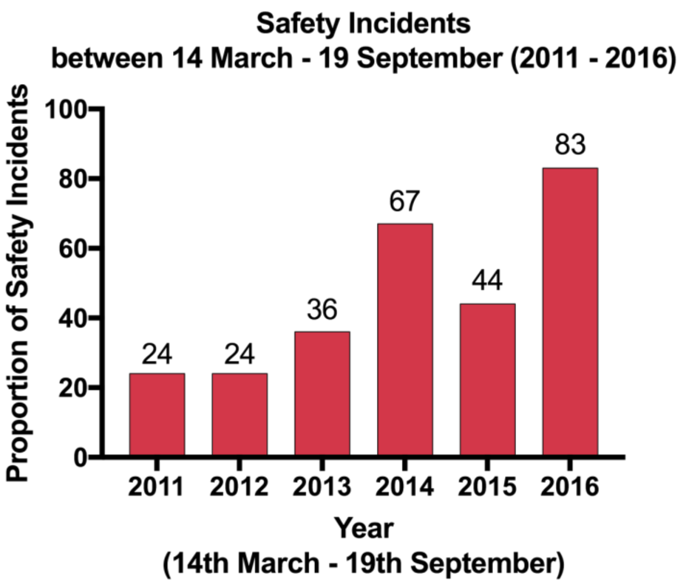


Figure 23: Total number of safety incidents during the period when train drivers were restricted from working their assigned rest days (2016) and the same historical period (2011 – 2015)

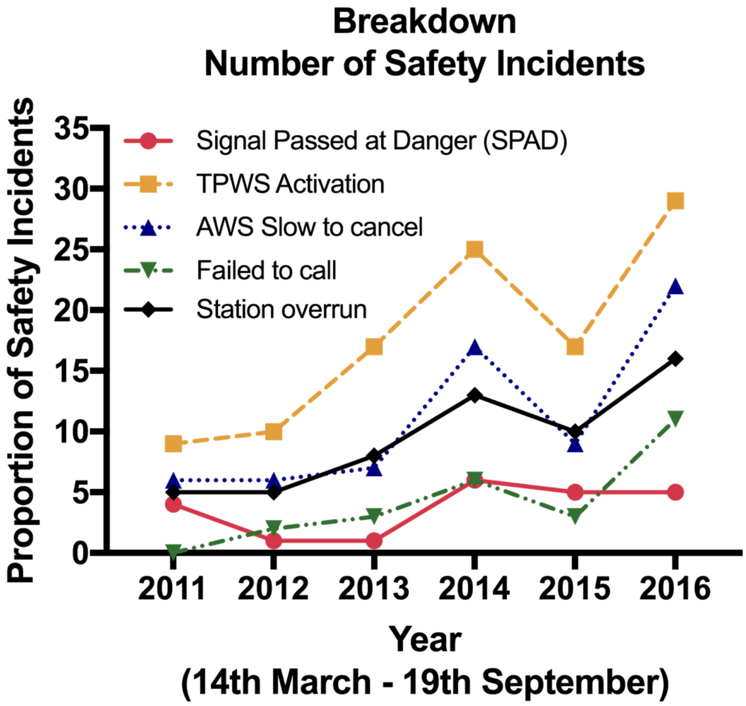


Figure 24: A breakdown of the total number of safety incidents during the period when train drivers were restricted from working their assigned rest days (2016) and the same historical period (2011 – 2015)

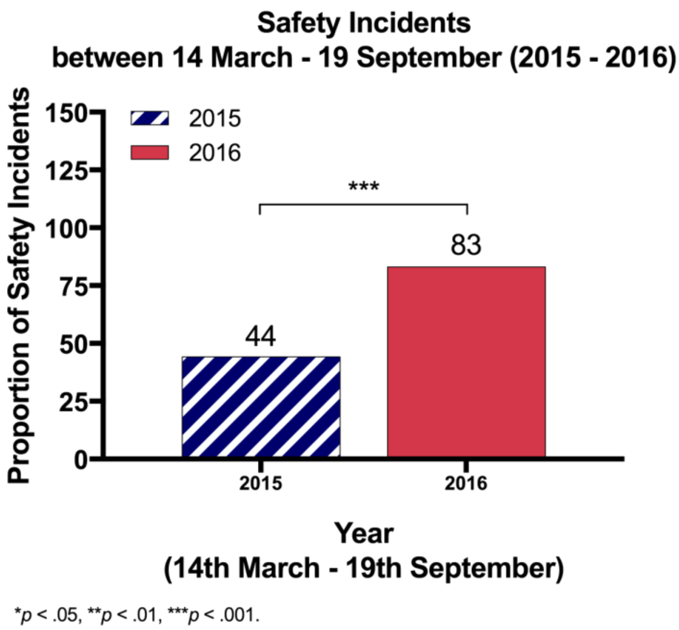


Figure 25: Total number of safety incidents when train drivers were restricted from working their assigned rest days (2016) and the previous year (2015)

3.4.2.1: Fatigue index scores and risk index scores across all six time periods

A Kruskal-Wallis Test was carried out to investigate whether there was a significant difference in the fatigue index (FI) scores of train drivers across the six time periods.

There was no statistical difference in the FI scores of train drivers across all six time periods, χ^2 (5, n = 278) = 2.03, p > .05 (see Figure 26).

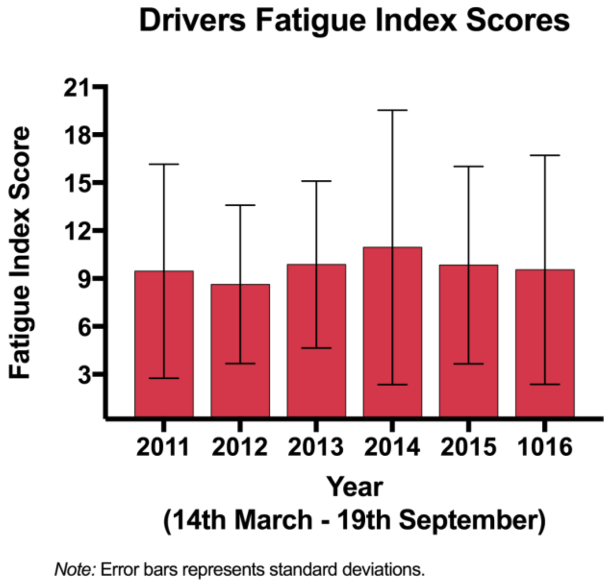


Figure 26: Fatigue index scores of train drivers across all six time periods

In addition, a Kruskal-Wallis Test was also carried out to investigate whether there was a significant difference in the risk index (RI) scores of train drivers across the six time periods. There was no statistical difference in the RI scores of train drivers across all six time periods, χ^2 (5, n = 278) = 5.64, p > .05 (see Figure 27).

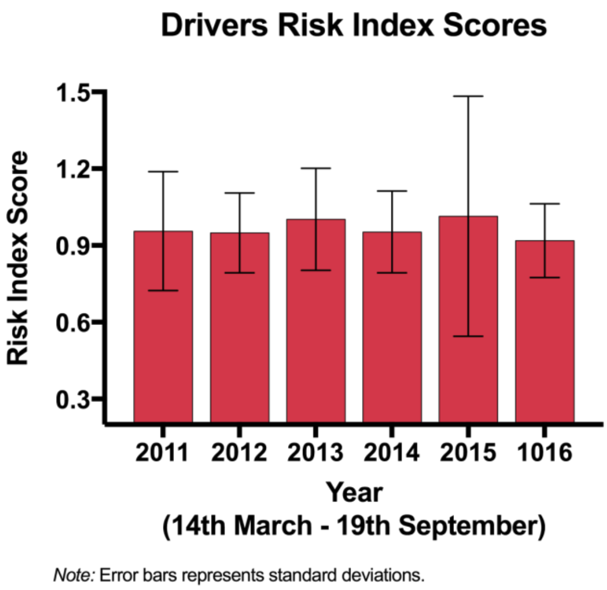


Figure 27: Risk index scores of train drivers across all six time periods

3.4.3: Discussion

The aim of this study was to investigate whether restricting train drivers from working on their assigned rest days based on a naturally occurring intervention reduced the number of safety incidents in which fatigue could have been a contributing factor i.e., Signal Passed at Danger (SPAD), Train Protection & Warning System (TPWS) activation, Automatic Warning System (AWS) slow to cancel, failed to call, and station overrun. In this study it was found that there was no statistical difference in safety incident frequencies throughout the six time periods. However, there was a 47 per cent statistically significant increase in the observed frequency of safety incidents between the time period of the naturally occurring intervention (i.e., 2016; n = 83) and the previous year (i.e., 2015; n = 44). These findings are not consistent with those found in the literature, whereby it has been identified that increased workload can have significant implications for safety incidents (Thomas, Paterson, Jay, Matthews, & Ferguson, 2019) or that high workload impact safety critical workers in their ability to localise audible alarm warnings i.e., in-attentional deafness (Edworthy *at al.*, 2018). The Automatic Warning System (AWS) audible indicator generates an audible warning alarm (i.e., a horn sound or electronic equivalent) (see RSSB, 2015b, for review), which must be acknowledged by the train driver within two to three seconds or an emergency brake application will occur i.e., AWS slow to cancel. However, it is important to point out that there are other influencing external factors. HSE (2006) identified that personal factors could contribute towards higher levels of fatigue, such as home-life balance as well as conflicts. In addition, RSSB (2012) presented other external factors, such as individual factors (e.g., lifestyle, age, diet, illness, medical conditions, drugs and alcohol use, etc.), and environmental factors both external to work (e.g., family circumstances, domestic responsibilities, sleep environment, etc.) and internal to work (e.g., unit condition, track condition, breakroom facilities, napping facilities, weather conditions, etc.) that could also contribute towards fatigue. It is important to highlight that these external factors are not integrated into the Health and Safety Executive (HSE) Fatigue Risk Index (FRI) calculator's biomathematical modelling (BMM), either on the short-term or cumulative fatigue component (Spencer *et al.*, 2006). However, discussions with various managers across Arriva Trains Wales (ATW) as well as train drivers revealed that during the period of the naturally occurring intervention, there were negotiations taking place between senior management and drivers company council (DCC) e.g., Associated Society of Locomotive Steam Enginemen and Firemen (ASLEF) and the National Union of Rail, Maritime and Transport (RMT), which fostered perceived uncertainties among train drivers as to whether any changes such as the restriction to work on rest days would remain a permanent amendment to the work arrangements for drivers (ATW, 2016). Research carried by Fan and Smith (2017a) at Arriva Trains Wales (ATW) found an association between perceived levels of stress at work and fatigue. Conversely, Mantler, Matejicek, Matheson and Anisman (2005) identified that coping with employment uncertainty was associated with higher levels of perceived stress, specially under low uncertainty conditions.

Analyses of the fatigue index (FI) and risk index (RI) scores revealed that there was no significant difference between the naturally occurring intervention and the same time period for each historical year. These findings seem to indicate that despite ATW restricting train drivers from working on their rest day, train drivers were still averaging very similar FI and RI scores as the same historical time periods. Discussions with train drivers at ATW revelated that during the naturally occurring intervention, train drivers who would have otherwise been rostered as a 'spare' or 'stand-by' were assigned/allocated a diagram or partial diagram (i.e., roster turn) to cover a vacant turn(s), which could also potentially extend their datum time (i.e., booking-on or booking-off time) (ATW, 2016). Under these circumstances, these findings could be explained through an association between increased workload and fatigue (Caldwell, Caldwell, Thompson, & Lieberman, 2019; Sadeghniiat-Haghighi & Yazdi, 2015). Fan & Smith (2019) identified that workload and working overtime were contributing factors towards fatigue within Arriva Trains Wales (ATW). However, Dorrian *et al.* (2011) highlighted that work duration and workload significantly influenced levels of fatigue, as well as an increased probability of a safety accident occurring as a result of the shift duration increasing (Goode, 2003). Alternatively, one cannot exclude financial circumstances whereby train drivers could have potentially extended their daily duty to compensate for loss in earnings. Mathisen and Bergh (2016) argued that working overtime increases error rates and rule violation.

Rumination in the cab could also be another potential explanation of these findings, especially when considering train drivers are not previewed to negotiation discussions between senior management and DCC e.g., ASLEF and RMT. As a result, train drivers would not know if these restrictions in working during rest days would become standard policies. Several researchers have outlined a relationship between changes in the workplace (e.g., uncertainties) and the increase in work-related stress (Karasek & Theorell, 1990; Schnall, Belkic, Landsbergis, & Baker, 2000). This is further supported by an association between perceived levels of stress and fatigue previously found at Arriva Trains Wales (ATW) (Fan & Smith, 2017a).

Since the HSE's FRI calculator is based on pre-defined parameters i.e., short-term fatigue; time of day, shift duration, rest period, and breaks as well as the build-up in fatigue that is associated with the continual disruption of sleep and the time required to recover from the period of sleep (Spencer *et al.*, 2006), the HSE's FRI calculator fails to take into account other fatigue-related external factors beyond the work-related parameters e.g., individual factors and environmental factors (RSSB, 2012). As a result, the HSE's FRI calculator does not factor in the various unique safety critical challenges train drivers experience which could be contributing factors towards fatigue. Therefore, there is a clear need to better understand the environmental factors that could contribute towards fatigue from within the cab environment. As a result, in-cab observations are a necessity to better understand the issues that could contribute towards safety incidents when fatigued from within the cab environment.

3.4.4: Study summary

The aim of this study was to investigate whether restricting train drivers from working on their assigned rest days based on a naturally occurring intervention reduced the number of safety incidents in which fatigue could have been a contributing factor. In this study, the observed frequency of safety incidents during the naturally occurring intervention time period was compared across the same dates in previous years to be able to determine whether restricting train drivers from working their assigned rest days significantly reduced the number of safety incidents in which fatigue could have been a contributing factor. Safety incident reports of 278 train drivers (266 men and 12 female) ranging in age from 24 – 63 years old (M = 47.46, SD = 7.48) were collected using the Safety Management Information System (SMIS). In this study it was found that there were no statistical differences in safety incident frequencies throughout the six time periods. However, there was a 47 per cent statistically significant increase in the observed frequency of safety incidents between the time period of the naturally occurring intervention (i.e., 2016; n = 83) and the previous year (i.e., 2015; n = 44). These findings seem to indicate that perhaps other external factors such as individual and environmental factors may also be contributors, which are not part of the biomathematical model within the Health and Safety Executive (HSE) Fatigue Risk Index (FRI) calculator. As a result, there is a clear need for better understanding some of the environmental factors that could contribute towards safety incidents in frontline safety critical workers, such as train drivers.

3.5: Chapter summary

The aim of this chapter was to investigate the effectiveness of the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator used at Arriva Trains Wales (ATW) as part of their Fatigue Risk Management System (FRMS) for monitoring and managing safety incidents in which fatigue could have been a contributing factor by answering two fundamental questions. The first study was carried out to investigate whether the present biomathematical model (BMM) for assessing train drivers' level of fatigue at Arriva Trains Wales (ATW) i.e., the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator related to the number of safety incidents in which fatigue could have been a contributing factor. In this study it was found that ATW's use of the HSE's FRI calculator as an objective measure of fatigue may not be an effective or accurate predictor of train driver's fatigue or risk levels, since analyses of 578 accessible safety incidents in which fatigue could have been a contributing factor only identified 0.2 per cent (n = 1) and 2.1 per cent (n = 12) of FI scores and RI scores respectively, that exceeded ATW's Fatigue Risk Management System (FRMS) recommended thresholds. As a result, ATW may need to carefully consider and evaluate with the appropriate urgency whether their current the HSE's FRI calculator's BMM is effective enough alone at predicting safety incidents in which fatigue could have been a contributing factor in frontline safety critical workers, such as train drivers. However, it found that Train Protection Warning System (TPWS) Activation was the most prominent safety incident for train drivers who had between 6 – 10 years of driving experience. Therefore, it is recommended that going forward, TfWRS' Safety, Training and Update Day (STUD) programme reflects and addresses these high TPWS Activation safety incidents. The second study was carried out to investigate whether restricting train drivers from working on their assigned rest days based on a naturally occurring intervention reduced the number of safety incidents in which fatigue could have been a contributing factor. In this study it was found that there was a 47 per cent statistically significant decrease in the observed frequency of safety incidents between the time period of the naturally occurring intervention (i.e., 2016; n = 83) and the previous year (i.e., 2015; n = 44). These findings seem to indicate that perhaps other external factors such as individual and environmental factors may also be contributors, which are not part of the biomathematical model within the HSE's FRI calculator. As a result, there is a clear need to identify some of the external environmental factors that could contribute towards safety incidents when fatigued through ethnographic research based on extensive in-cab observations.

Chapter 4: Pre-unit mobilisation procedures and checks as well as in-cab observations

4.1: Overview of chapter

The previous chapter outlined the current biomathematical model (BMM) used at Arriva Trains Wales (ATW) as part of their Fatigue Risk Management System (FRMS), i.e., the Health and Safety Executive’s (HSE) Fatigue Risk Index (FRI) calculator. The previous chapter summarises that the HSE’s FRI calculator used by ATW was not effective at predicting safety incidents in which fatigue could have been a contributing factor in frontline safety critical workers, such as train drivers. In addition, it was identified that restricting train drivers from working on their assigned rest days based on a naturally occurring intervention did not reduce the number of safety incidents in which fatigue could have been a contributing factor. Therefore, there is a clear need to better understand some of the environmental factors that could contribute towards safety incidents at ATW. As a result, the aim of this chapter is to identify some of the external environmental factors that could contribute towards safety incidents when fatigued, through ethnographic research based on extensive in-cab observations.

4.2: Introduction and rationale

The main focus of the ethnographic qualitative research approach is to provide rich, holistic insights into people’s views and actions (Reeves, Kuper, & Hodges, 2008). Dixon-Woods and Bosk (2010) state that ethnography is the direct observation of people. Therefore, for over 35 years ethnography has been defined as the use of participant observation (Holy, 1984). Reeves *et al*. (2008) state that the use of participant observation enablse ethnographers to be able to immerse themselves in a situation or location that promotes deeper understanding as well as its subtleties in different contexts, which are usually hidden from the public literature. However, Savage (2000a) suggests that ethnography is considered both as contextual and reflexive, which emphasises the significance of being able to apply context in order to better understand events and meaning, while also taking into account the impact of the researcher's presence as well as the research strategy on the observations e.g., the train cab's environment, surrounding sounds, visual stimuli, etc.

Dixon-Woods and Bosk (2010) identify that ethnographers should seek to firstly provide an empirical description of what happens and secondly to produce an analysis of what they see, which is quite specific to the observer's setting. However, according to Savage (2000a), overall consensus on the theory of knowledge that underpins an ethnographic account among ethnographers is yet to be reached. As a result, the task of an ethnographer is to record the culture, the perspectives and practices of the people in these settings by seeing the world from their point-of-view (Reeves *et al*., 2008). However, Savage (2000b) points out that ethnography entails prolonged fieldwork. According to Leslie, Paradis, Gropper, Reeves, and Kitto (2014), ethnography involves the direct observation of participant's behaviour within their environment, which is always over a sustained period of time, for example, the observation of a train driver within the cab environment for at least the length of a diagram turn. However, since ethnographers invest a significant duration of time observing actions, it can sometimes be quite challenging to secure frequent access, especially if the organisation has concerns that the observations may cast them in an unflattering perspective (Reeves *et al*., 2008). Nevertheless, ATW granted the researcher unrestricted access to various safety systems, including frontline safety critical train drivers. Only one similar study was found that explored the isolated roles and responsibilities of London Underground train drivers (see Heath, Hindmarsh, & Luff 1999, for review). Since then, there has been no published ethnographic research carried out from within the cab environment to determine some of the issues that train drivers experience that could potentially contribute towards fatigue. Therefore, the aim of this chapter is to identify some of the external environmental factors that could contribute towards safety incidents when fatigued through ethnographic research based on extensive in-cab observations.

4.3: Methodology

The researcher fully immersed himself and completed ~120 hours of consecutive in-cab observations spilt between Valleys & Cardiff Local Routes (formally known as Cardiff Valley lines) and Cardiff Mainline Routes. This was achieved through the observer accompanying four train driving instructors who held a Category B train driving licence under the Train Driving Licences and Certificates Regulations 2010 (TDLCR).

4.3.1: Ethical approval

The study received ethics approval from Cardiff University's Ethics Committee (EC.17.09.12.4947R2A). The study conformed to the seventh amendment of the Declaration of Helsinki 1964 (World Medical Association, 2013) and was in accordance with the Data Protection Act 1998 and the General Data Protection Regulation (GDPR). Cab observations were cleared by Arriva Trains Wales (ATW) attending authorities (e.g. Head of Drivers, Head of Safety, and Director of Operations) before ATW authorised and issued an Arriva Train Wales Driving Cab Pass. All observations were undertaken in accordance with the BPS (2014: 25) Code of Human Research Ethics guidelines concerning observation in a natural setting as well as the observer adhering to all instructions and commands as directed by the train driving instructor to ensure safety was never compromised as a result of distractions.

4.3.2: Participants

The observer wanted to ensure the highest external reliability and validity of the ethnographic research (see LeCompte & Goetz, 1982, for review) of the in-cab observation. Therefore, in order for the observer to be in a position to be able to discover the same phenomena as well as the ability to generate the same constructs from within the cab setting, the observer carried out multiple roster observations that also included observations of multiple train drivers. The observer was paired up with four experienced train driving instructors who were highly competent in operating trains for the carriage of passengers as well as being able to carry out their duties and responsibilities while being accompanied by a third-party individual inside the cab i.e., the observer. Two of the four train driving instructors were from the Valleys & Cardiff Local Routes and the other two were from the Cardiff Mainline Routes. Observations lasted for the total duration of four consecutive weeks, which represented ~120 hours of in-cab observation (see Table 3), both split into early shifts i.e., booking-on between 03:41 – 09:55 and late shifts i.e., booking-off between 15:55 – 00:47 (see Table 4, for a breakdown). Due to the limited number of train driving instructors on the Valleys & Cardiff Local Routes and on the Cardiff Mainline, demographic characteristics of all four train driving instructors are not reported to ensure full anonymity in accordance with the Data Protection Act 1998 and the General Data Protection Regulation (GDPR).

Train Driver Instructors	Total Combined Diagram Turns (hh:mm)
1 (Valleys & Cardiff Local Routes)	35:34
2 (Valleys & Cardiff Local Routes)	32:27
3 (Cardiff Mainline Routes)	36:31
4 (Cardiff Mainline Routes)	14:37
Total	**119:09**

Table 3: Total combined diagram turns

Train Driver Instructors	Book On	Book Off	Total Hours	Diagram	Total Diagram Turn (hh:mm)
1 (Valleys & Cardiff Local Routes)	08:48	18:00	09:12	424	09:12
	09:48	18:29	08:41	463	08:41
	09:55	18:24	08:29	428	08:29
	09:49	19:01	09:12	425	09:12
2 (Valleys & Cardiff Local Routes)	05:59	15:15	09:16	403	09:16
	08:30	16:00	07:30	TB (Team Brief)	07:30
	05:50	14:43	08:53	414	08:53
	05:41	12:29	06:48	2411	06:48
3 (Cardiff Mainline Routes)	15:55	00:34	08:39	385	08:39
	15:19	00:47	09:28	315	09:28
	15:19	00:47	09:28	315	09:28
	15:46	00:42	08:56	2215	08:56
4 (Cardiff Mainline Routes)	05:30	11:40	06:10	230	06:10
	05:50	14:17	08:27	403	08:27
	-	-	-	RD (Rest Day)	-
	-	-	-	RD (Rest Day)	-

Table 4: Breakdown of cab observation diagram turns

4.3.3: Design and procedure

Pink and Morgan (2013) outlines that there are two types of ethnography; long-term and short-term. Long-term ethnography is a methodology whereby the observer becomes an apprentice through the process of developing new skills with the goal to becoming a practitioner e.g., acquiring all the relevant rules and regulations knowledge, traction knowledge, route knowledge, and practical train handling to be able to become a train driver. However, the commitment involved to become a certified train driver was far beyond the duration and scope of the doctoral research framework as well as the feasibility in terms of Arriva Train Wales' (ATW) resources for the observer to obtain all the relevant competent areas to be able to hold a Category B train driving licence under the Train Driving Licences and Certificates Regulations 2010 (TDLCR). Therefore, short-term ethnography was felt to be the most appropriate methodology. According to Pink and Morgan (2013), short-term ethnography consists of developing different possibilities for engagement with the knowledge of other individual's experiences e.g., learning from train driving instructors who have extensive rules and regulations knowledge, traction knowledge, route knowledge, and practical train handling experience as well as the safety culture of both the rail industry and ATW.

The direct observations of four train driving instructors were made from within the cab during normal operations for the duration of four consecutive weeks. All observations were based on interactions solely between the train driving instructor and the researcher, which resulted in a strong relationship that allowed for an open channel of dialogue that would otherwise not have been possible if the observer had not spent a significant duration in the cab (i.e., on average 30 hours with each train driving instructor), building mutual trust and respect, or if other ATW employees were present in the cab e.g., train driver managers. The observer plugged into the train driving instructors' full weekly roster, which was comprised of various diagram turns, providing the observer with an opportunity to be exposed to a variety and wide selection of potential situations that could result in lapse in attention or situation awareness, which could contribute towards safety incidents when feeling the ill effects of fatigue.

The researcher fully immersed himself and completed ~120 hours of consecutive in-cab observations. Through the successful pairing by the Head of Drivers between the observer and four train driving instructors, the observer was able to observe, discuss and reflect on the duties and responsibilities of the train driving instructors during normal operations as well as obtaining fundamental rules and regulations knowledge, traction knowledge, and route knowledge, which also included a comprehensive understanding of the practical train handling that were all obtained and experienced within the cab. Through the process of full immersion, the observer was able to gain a unique perspective into the various issues train drivers encounter and experience that could contribute towards safety incidents when feeling the ill effects of fatigue.

4.4: Cab observations

The aim of this study was to identify some of the external environmental factors that could contribute towards safety incidents when fatigued through ethnographic research based on extensive in-cab observations. Therefore, these findings are based on the researcher's own unique observational perspective of frontline safety critical train driver.

It is important to point out the observer should be fully compliant with all rules and traction guidelines, since all observations took place from within the cab of an operational trains that were transporting passengers. The observer strictly adhered to all Arriva Train Wales (ATW), Rail Safety and Standards Board (RSSB) and Office of Rail and Road (ORR) rules required for train drivers during turn of duty. Below is a comprehensive list of all the information the observer had to familiarise himself with prior to commencing cab observations. These were:

- Professional Driving Policy (ATW, 2017b);
- Arriva Trains Wales working arrangements for drivers (ATW, 2016);
- Rule Book: Train Driver Manual (RSSB, 2015b; RSSB, 2018b);
- Alcohol and Drugs Policy (ATW, 2017c);
- Company Safety Policy (ATW, 2017c);
- Personal Protective Equipment (PPE) (ATW, 2017c);
- Token Protocol (RSSB, 2014a; RSSB, 2016a);
- Low adhesion (ATW, 2017b; RSSB, 2015c);
- Relevant Arriva Train Wales publications:
    - Daily roster appearance sheet
    - Diagram schedule sheet
    - Signing-on sheet
    - SPAD notice case board
    - Weekly operating notice booklet
    - Late notice case board
    - New notice case board
    - General notice case board
    - Permanent notice case board
    - Seasonal notice case board
    - Low adhesion board
    - Health and safety notice case board
    - ASLEF notice board
- On-Train Monitoring Recorder (OTMR; RSSB, 2014b);
- Global System for Mobile Communications – Railway (GSM–R; RSSB, 2015d; RSSB, 2017a; RSSB, 2018c);
- Driver Advisory Systems (DAS; RSSB, 2009; RSSB, 2010b);
- Driver Reminder Appliance (DRA; McCorquodale, 2002);
- Personal Track Safety (PTS; RSSB, 2015e).

4.5: Lifestyle and preparation before shift

To ensure full immerse, two days prior to commencing observations, the observer ensured he was well rested and fully complied with all regulations, such as ensuring the observer adhered to the Alcohol and Drugs Policy (ATW, 2017c) as well as fit to drive/work trains (i.e., fitness for duty). Therefore, the night before the turn, the observer ensured he had all the right Personal Protective Equipment (PPE; ATW, 2017c), such as; steel toe cap shoes, hi vis vest, safety glasses, ear plugs, and a pair of correction spectacles, and all meals prepared in the event that either the Main Rest Break (MRB) and/or Short Rest Break (SRB) were assigned in locations that has fewer options to purchase food such as at Cardiff Shed Depot (CSD), which has no nearby facilities to purchase warm food. It was also vital for the observer to review the turn on the daily roster appearance sheet the day before it was due to commence to best prepare for any unforeseeable amendments to the roster turn. Moreover, the observer found it useful to gain a much deeper understanding of the route learning process prior to commencing the observation, which involved the observer spending valuable time in the manager resources office to better understand signalling (see Figure 28, for a visual representation) as well as the train simulator for traction and route knowledge (see Figure 29, for a visual representation) to acquire an insight as well as an understanding of the various procedures that are implemented, for example, when a train is delayed or cancelled due to operational/equipment failure or safety related incidents.

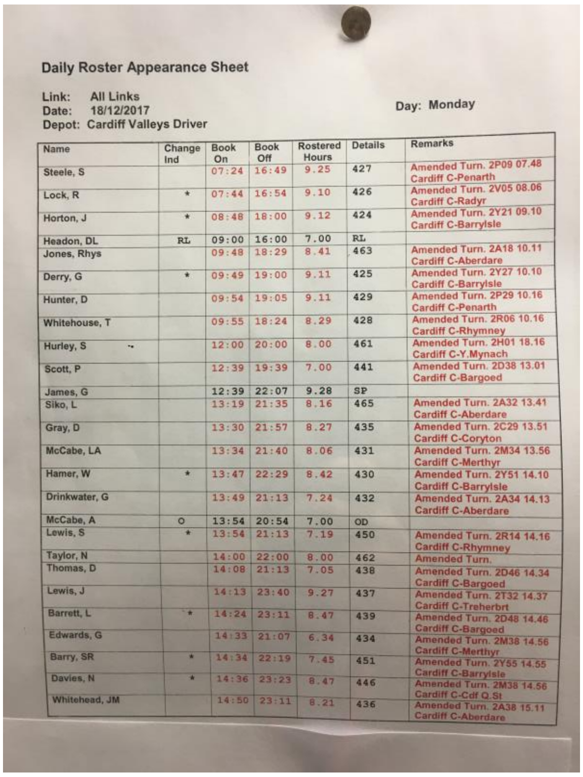
Figure 28: Observer in the manager resources office learning basic signalling

Figure 29: Observer in the train simulator acquiring basic traction and route knowledge

4.6: Daily roster appearance sheet

The main purpose of the daily roster appearance sheet is to inform train drivers of any amendments to their turn. Since the rostered turn could be modified + or – 1 hour from the original time i.e., diagram amendment from long-term planning (LTP) to short-term planning (STP) up to 24 hours prior to the booking on time whether planned or unplanned (ATW, 2016), train drivers must remain flexible and accommodating to amendments, as well as ensure they have reviewed the latest version of the daily roster appearance sheet (see Figure 30 and Figure 31, for a visual representation). Therefore, the observer ensured he reviewed and confirmed the upcoming turn from the daily roster appearance sheet, and had revised the book-on, book-off and any amendments to the turn. This extra step ensured a much-needed validation and reassurance of the booking-on time as the observer was only provided with the train driver instructor's name, diagram number as well as the turn's book-on and book-off times several days prior to the observation commencing, which could have potentially been amended. The daily roster appearance sheet also served to verify and confirm any changes to upcoming rostered turns.

Figure 30: Daily roster appearance sheet board



Daily Roster Appearance Sheet

Link: All Links
Date: 18/12/2017
Depot: Cardiff Valleys Driver

Day: Monday

Name	Change Ind	Book On	Book Off	Rostered Hours	Details	Remarks
Steele, S		07:24	16:49	9.25	427	Amended Turn. 2P09 07.48 Cardiff C-Penarth
Lock, R	*	07:44	16:54	9.10	426	Amended Turn. 2V05 08.06 Cardiff C-Radyr
Horton, J	*	08:48	18:00	9.12	424	Amended Turn. 2Y21 09.10 Cardiff C-Barrylsle
Headon, DL	RL	09:00	16:00	7.00	RL	
Jones, Rhys		09:48	18:29	8.41	463	Amended Turn. 2A18 10.11 Cardiff C-Aberdare
Derry, G	*	09:49	19:00	9.11	425	Amended Turn. 2Y27 10.10 Cardiff C-Barrylsle
Hunter, D		09:54	19:05	9.11	429	Amended Turn. 2P29 10.16 Cardiff C-Penarth
Whitehouse, T		09:55	18:24	8.29	428	Amended Turn. 2R06 10.16 Cardiff C-Rhymney
Hurley, S		12:00	20:00	8.00	461	Amended Turn. 2H01 18.16 Cardiff C-Y.Mynach
Scott, P		12:39	19:39	7.00	441	Amended Turn. 2D38 13.01 Cardiff C-Bargoed
James, G		12:39	22:07	9.28	SP	
Siko, L		13:19	21:35	8.16	465	Amended Turn. 2A32 13.41 Cardiff C-Aberdare
Gray, D		13:30	21:57	8.27	435	Amended Turn. 2C29 13.51 Cardiff C-Coryton
McCabe, LA		13:34	21:40	8.06	431	Amended Turn. 2M34 13.56 Cardiff C-Merthyr
Hamer, W	*	13:47	22:29	8.42	430	Amended Turn. 2Y51 14.10 Cardiff C-Barrylsle
Drinkwater, G		13:49	21:13	7.24	432	Amended Turn. 2A34 14.13 Cardiff C-Aberdare
McCabe, A	O	13:54	20:54	7.00	OD	
Lewis, S	*	13:54	21:13	7.19	450	Amended Turn. 2R14 14.16 Cardiff C-Rhymney
Taylor, N		14:00	22:00	8.00	462	Amended Turn.
Thomas, D		14:08	21:13	7.05	438	Amended Turn. 2D46 14.34 Cardiff C-Bargoed
Lewis, J		14:13	23:40	9.27	437	Amended Turn. 2T32 14.37 Cardiff C-Treherbrt
Barrett, L	*	14:24	23:11	8.47	439	Amended Turn. 2D48 14.46 Cardiff C-Bargoed
Edwards, G		14:33	21:07	6.34	434	Amended Turn. 2M38 14.56 Cardiff C-Merthyr
Barry, SR	*	14:34	22:19	7.45	451	Amended Turn. 2Y55 14.55 Cardiff C-Barrylsle
Davies, N	*	14:36	23:23	8.47	446	Amended Turn. 2M38 14.56 Cardiff C-Cdf Q.St
Whitehead, JM		14:50	23:11	8.21	436	Amended Turn. 2A38 15.11 Cardiff C-Aberdare

Figure 31: Close up of a typical daily roster appearance sheet

4.7: Responsibilities before driving

At the starting stage of the observation there was a large volume of information to absorb, digest and comprehend. However, over time the focus shifted from global information processing to selective attention to informational changes (i.e., local information processing; Lewis, Ellemberg, Maurer, Dirks, Wilkinson, & Wilson, 2004; Neiworth, Gleichman, Olinick, & Lamp, 2006) in the form of absorbing, digesting and comprehending only the latest displayed information and/or cases that were being added to the various notice boards. Below is a list of all the various notice boards and cases that all train drivers must read in no particular order before commencing their rostered diagram duties. It is good practice for all train drivers to adopt a preferred order sequence and strictly adhere to that said order so that nothing is ever overlooked or missed out. The observer adopted a preferred order of sequences that he felt made the most logical sense to him, which is presented below. The advantage of adhering to a pre-defined order of sequences is that in the event that attention is drawn away from the displayed information e.g., talking to other drivers, guards, resources managers, etc., the train driver is able to quickly return to the exact step prior to the interruption by repeating their preferred order of sequences while acknowledging those that had already been completed, making it far easier to know what is left to view as well as ensuring nothing is overlooked or missed out that could potentially be vital knowledge while driving the train on a particular unit type or route.

4.7.1: Fitness for duty visual inspection

The booking-on office for the observer was Cardiff Central. At Cardiff Central, the resources managers team visually inspects train drivers to ensure fitness for duty when they come to collect their duty diagram (see Figure 32). Regulation 24 of the Railways and Other Guided Transport Systems (Safety) Regulations 2006 (as amended; ROGS) requires that controller of safety critical work ensure that staff are competent and fit to undertake their safety critical activities. In addition, regulation 25 of ROGS requires that safety critical workers such as train drivers are not so fatigued that the health and safety of the train driver or of other persons on the transport system could be significantly affected. However, other depots within ATW adopt an offsite booking-on systems, which means that not all booking-on locations across the franchise carry out a visual inspection of fitness for duty. Moreover, other Train Operating Companies (TOC) implement automated booking-on systems, which bypasses human interaction completely unless there have been any amendments to the drive's duty diagram, which would then require the train driver to speak to an offsite resources manager representative. The visual inspection is usually carried out by the resources team, who would observe train drivers at the booking-on office using their best judgement to visibly inspect individuals, and determine whether there are any indicators that the train driver would be prevented from carrying out their duties safely, such as indicators atypical of the train driver's behaviour e.g., smell of alcohol, stumbling or tripping, speech impairment, not having their corrective glasses, etc. but to name a few.

Figure 32: Observer fitness for duty

4.7.2: Diagram schedule sheet

The diagram schedule sheet provides a breakdown for the turn the train driver has been assigned to operate. This will include all the relevant information the driver will need for their duty roster, which includes; the diagram number; booking-on, booking-off and the unit type (e.g., Class 143, Class 150, etc.) Figure 33 and Figure 34 provide an example of the typical diagram schedule sheet a train driver would be assigned at Arriva Trains Wales (ATW). It is important to note that it is the responsibility of the train drivers to ensure they are fully component to operate all unit types (e.g., Class 140s, Class 150s, etc.) presented in the diagram schedule sheet as well as having all the required route knowledge to operate their assigned diagram schedule sheet.

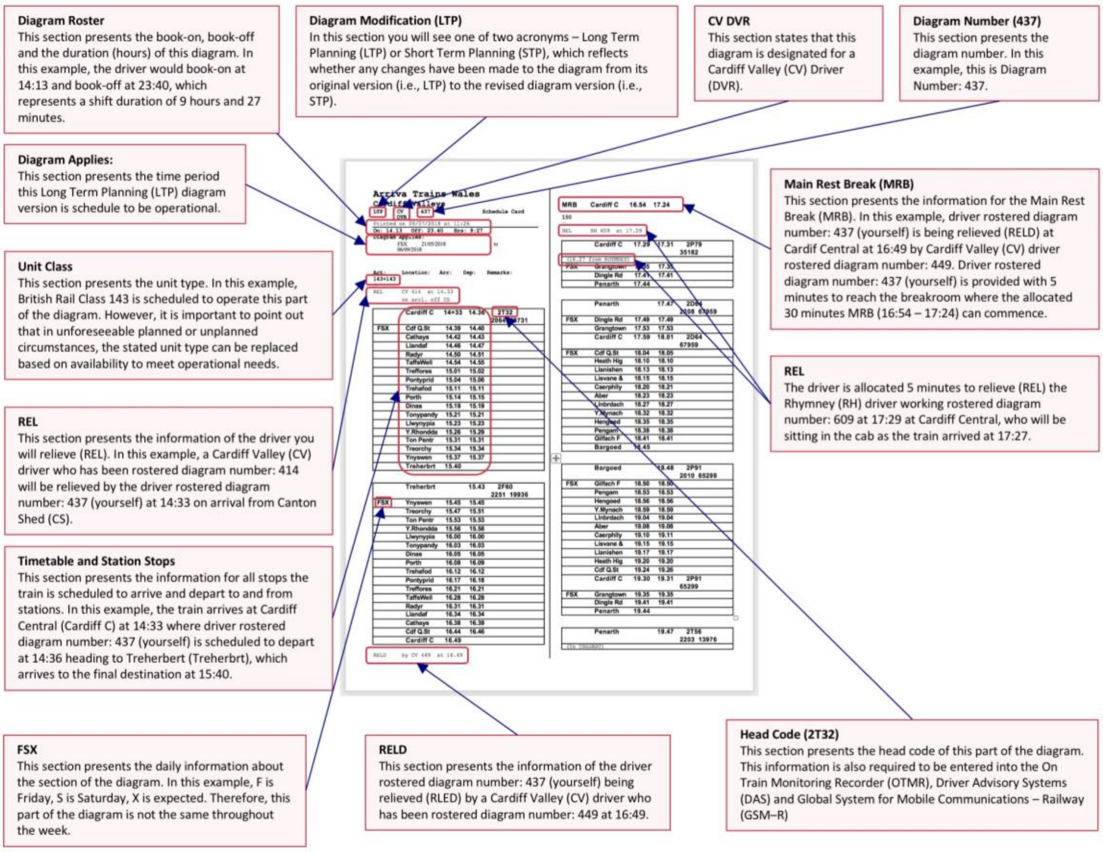


Figure 33: Diagram breakdown (page 1 of 2)

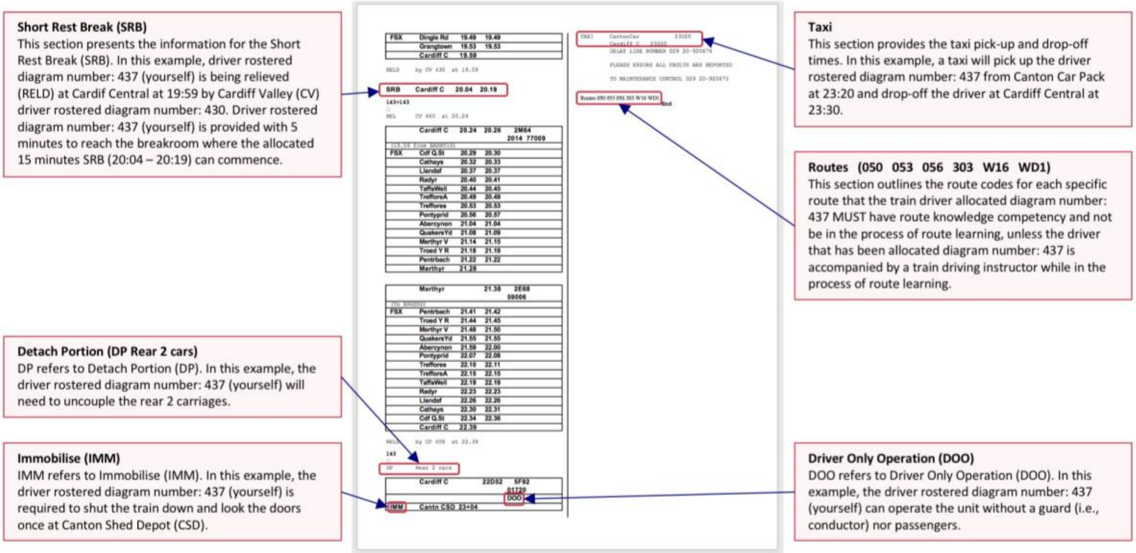


Figure 34: Diagram breakdown (page 2 of 2)

4.7.3: Signing-on sheet

The main purpose of the signing-on sheet (see Figure 35, for a visual representation) is for all train drivers to acknowledge that they have read, understood and will fully adhere to, the booking on statement (ATW, 2016) as well as to declare they have had the minimum rest period before signing on, which further validates that they are 'fit for duty'. In addition, the signing on sheet requires that train drivers write down their allocated diagram turn number, which further ensures that they are fully competent with the route and traction knowledge requirements to operate their assigned diagram schedule sheet (see RSSB, 2017b; RSSB, 2018d, for reviews).

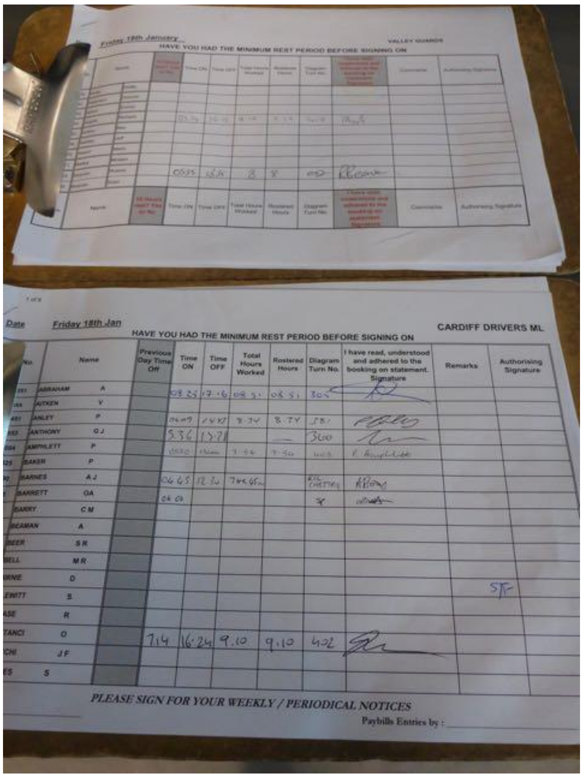


Figure 35: Signing-on sheet

4.7.4: SPAD notice case board

The Signal Passed at Danger (SPAD) notice case board (see Figure 36, for a visual representation) displays all the recent anonymous SPAD's that have occurred across the franchise as well as the route (i.e., location) of the SPAD, which includes images of the signal that the train driver had SPAD. In addition, the SPAD notice case board also provides train drivers with information of the various multi-SPAD signals across Wales, which serves as a reminder to train drivers to remain vigilant and alert at all times, with extra focus on a multi-SPAD signal. According to Moor (2013), a multi-SPAD signal refers to if there have been two or more SPADs at the signal during the five years prior to that point in time. Moreover, the SPAD notice case board also displays other Train Operating Company's (TOC) SPADs that operate on the same signal sections as ATW. It is important to point out that despite train drivers who are involved in SPAD incidents being anonymous, identifying who was involved in a SPAD is not too difficult as a train driver can easily extrapolate information from the SPAD alert case, such as date, time and route to determine who was working the assigned diagram when it is cross-referenced with the daily roster appearance sheet. Moreover, alternative strategies include train drivers who may have been otherwise assigned as a spare driver would see themselves catching (i.e., assigned) a diagram number, which was previously assigned to another driver on the daily roster appearance sheet, but that driver was no longer available to work that diagram turn.

Figure 36: Signal Passed at Danger (SPAD) notice case board

4.7.5: Weekly operating notice booklet

The weekly operating notice booklet (see Figure 37, for a visual representation) provides train drivers with all the planned weekly safety notices, temporary speed restrictions, engineering arrangements, signalling and permanent way alternative information. Train drivers must read, understand and fully adhere to what has been clearly outlined in great depth in the weekly operating notice booklet. However, there will be occasions when there are unplanned operational safety notices, temporary speed restrictions, engineering arrangements, signalling and permanent way alternative, which train drivers must be made aware of and which are displayed on the late notice case board (see Section 4.7.6).

PRIVATE AND NOT FOR PUBLICATION

Page 1 of 27

ATW
Cardiff Valleys

Week No.
37

WEEKLY OPERATING NOTICE

Containing

SAFETY NOTICES
TEMPORARY SPEED RESTRICTIONS
ENGINEERING ARRANGEMENTS
SIGNALLING AND PERMANENT WAY ALTERATIONS
(NEW ITEMS)
(DETAILS OF WORK ALREADY CARRIED OUT)
GENERAL INSTRUCTIONS AND NOTICES
(AMENDMENTS TO OPERATING PUBLICATIONS)
(MISCELLANEOUS INSTRUCTIONS)

SATURDAY 09 DECEMBER 2017
to
FRIDAY 15 DECEMBER 2017

THIS NOTICE COVERS SELECTED LINES ONLY

Printed on Fri Dec 01 2017
If the weekly supply of this notice is not received at the normal time, application must be made to your local Manager

Figure 37: Weekly operating notice booklet

4.7.6: Late notice case board

The late notice case board (see Figure 38, for a visual representation) provides train drivers with the most recent unplanned operational safety notices, temporary speed restrictions, engineering arrangements, signalling and permanent way alternative, which train drivers must be made aware of before commencing their scheduled diagram turn. An example of the kind of information that would be displayed on the late notice case board may include emergency unplanned speed restrictions due to unforeseeable / unplanned Network Rail engineering work. It is important to highlight that all planned operational safety notices, temporary speed restrictions, engineering arrangements, signalling and permanent way alternative are published in the weekly operating notice booklet as presented in Section 4.7.5.

Figure 38: Late notice case board

4.7.7: New notice case board

The new notice case board (see Figure 39, for a visual representation) provide train drivers with information from notices being transferred from the late notice case board.

Figure 39: New notice case board

4.7.8: General notice case board

The general notice case board provide train drivers with information that is non-safety critical in nature (i.e., traction or operational information). For example, a number of train drivers have been late, please check your booking-on time. Another example (see Figure 40, for a visual representation) can be seen below;

> *'This morning, an Arriva staff member was hit by a car outside No1 Central Square whilst being stood on the pavement after crossing the road. A car illegally mounted the kerb to drop off four workers for the building site and hit the staff member on his left side.*
>
> *We've not got a definitive date when the work will be complete, so could you please be extra vigilant in this area during this time. If you see any illegal traffic activity in this (SIC) areas, could you please report it, with a photo to your team manager.'*

Figure 40: General notice case board

4.7.9: Permanent notice case board

The permanent notice case board (see Figure 41, for a visual representation) provides train drivers with information such as safety statements that remain posted indefinitely as a continuous reminder that all train drivers must adhere to these announcements at all times.

Figure 41: Permanent notice case board

4.7.10: Seasonal notice case board

The seasonal notice case board (see Figure 42, for a visual representation) provides train drivers with information that directly affects the traction, track or signal. For example, when a section of track is flooded, the train driver needs to adhere to rules outlined by Arriva Train Wales (ATW; see Figure 43, for a visual representation). If water is above the top of the railhead, movement of the unit is only permitted by instructions from Operations Control. In addition, if the water is moving and likely to dislodge the ballast or has dislodged ballast, the movement of the unit needs to come to a complete stop and wait for further instructions from Operational Control. However, if the water is at any level of the railhead, movement is restricted to a maximum speed of 5 mph (10km/h).

Water levels below the bottom of the railhead are acceptable for normal working movement.

Figure 42: Seasonal notice case board

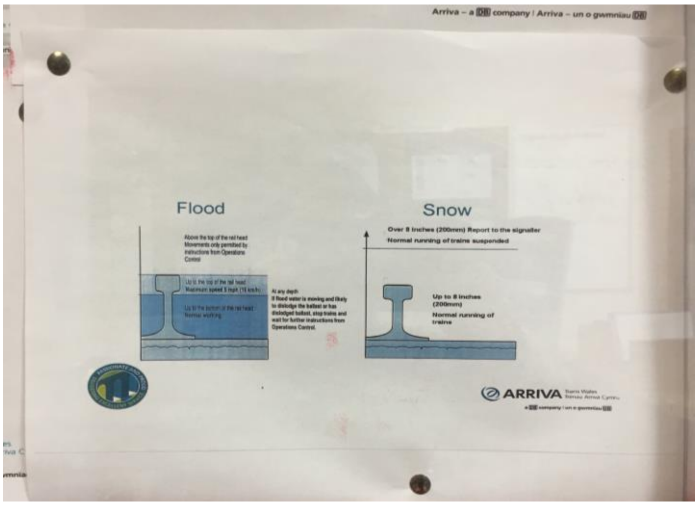


Figure 43: Example of a seasonal notice

4.7.11: Low adhesion board

The low adhesion board (see Figure 44, for a visual representation) provides train drivers with information on sections of the route that have low adhesion, i.e., grip (e.g., leaf contamination on the railhead) or how the sanders operate on different units.

Figure 44: Low adhesion board

4.7.12: Health and safety notice case board

The health and safety notice case board (see Figure 45, for a visual representation) provides train drivers with mandatory health and safety information that must be followed at all times. For example, at Canton train drivers have to wear safety glasses eye protection at the depot once they leave the carriage. In other words, before the train driver is able to put their foot on the depot, safety glasses eye protection must be put on.

Figure 45: Health and safety notice case board

4.7.13: ASLEF notice case board

ASLEF (Associated Society of Locomotive Engineers and Firemen) is Britain's trade union for train drivers. Therefore, the ASLEF notice case board (see Figure 46, for a visual representation) provides train drivers with information on the next branch meeting as well as important published information updating train drivers of various union related matters.

Figure 46: ASLEF notice case board

4.8: Setting up the cab environment

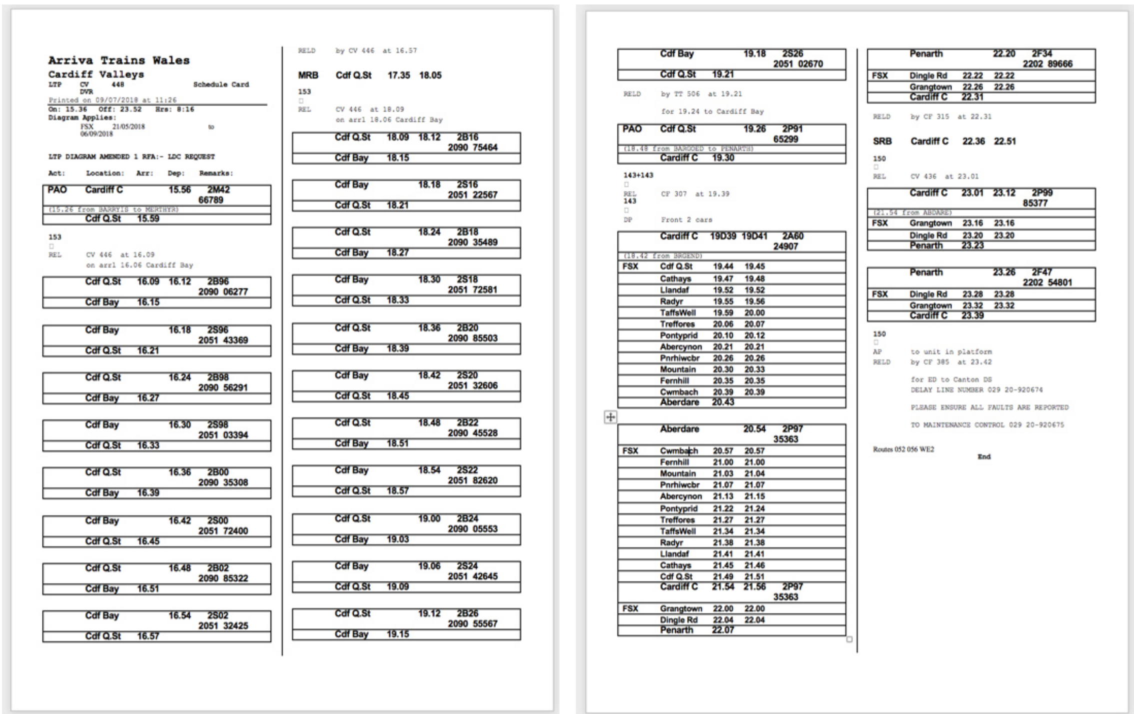
The observer experienced setting up the cab environment a significant number of times throughout the four weeks period as the observer was present during the procedure of a train driver being relieved (RELD), setting up the various logging-on in-cabin systems e.g., the Global System for Mobile Communications – Railway (GSM-R), the On-Train Monitoring Recorder (OTMR), and the Driver Advisory Systems (DAS). In addition, the procedures for changing unit ends (see Section 4.8.3) as well as the for preparing and moving a unit from Cardiff Shed Depot (CSD) was observed (see Section 4.8.4).

4.8.1: Visual inspection of the unit

One of the protocols within the cab environment involves the train drivers ensuring that the unit they are about to operate displays the final destination as well as ensuring the correct unit headlights are switched on (e.g., white for the front end and red for rear end). It was observed that as the unit approached the platform, the train driver visually inspected the final destination being displayed on the unit, which in some cases were correct when relieving a train driver on a partially completed route e.g., the unit route did not terminate at Cardiff Central and continued on to its terminal station platform. If this was not the case, the train driver would then be required to amend the destination blind to reflect the final destination of the section route as outlined on the diagram. Good observed practice was for the train driver to communicate with the train guard to verify the tail end of the train also reflected the terminal station platform of the unit. It was observed that establishing a friendly relationship facilitated in guards assisting by setting up the destination blind on approach to the terminal station platform. There seemed to be a mutual understanding and facilitation of this additional task by the Valley line guards but was not always the case on the main line. Despite this task being relatively simple and straight forward to execute, this task is done manually and requires the rolling through the various terminal stations until the correct station is displayed, which adds an extra minute in setting up the cab environment. Other train operating companies (TOCs) have implemented digital boards to display the terminal stations. Perhaps Arriva Train Wales (ATW) could introduce digital boards as part of their proposed new rolling stock (i.e., train units) strategy.

4.8.2: In-cabin systems

There are several in-cabin systems that the train driver needs to setup and log-in. For the purpose of simplicity, the observer will only describe the systems that require the train driver to log-in their credentials, which are; the On-Train Monitoring Recorder (OTMR), the Global System for Mobile Communications – Railway (GSM-R), and the Driver Advisory Systems (DAS). These in-cabin systems do follow a specific safety priority, but which system the train driver chooses to log-in first or the particular order can be carried out at their own discretion. However, train drivers tend to adopt a preferred order of sequences for their log-in procedures, which are for the same reasons that have been previously identified i.e., reducing human errors due to distractions or streamlining setup procedure due to continuous monotonous repetition (e.g., 10+ Cardiff Queen Street – Cardiff Bay return as it is a single track with one stop at each end), such as those that can be found at the start of a diagram (e.g., diagram number: 448, see Figure 47) or at end of a diagram (e.g., diagram number: 433, see Figure 48).



Arriva Trains Wales
Cardiff Valleys
LTP CV 448 DVR Schedule Card
Printed on 09/07/2018 at 11:26
On: 15.36 Off: 23.52 Hrs: 8:16
Diagram Applies:
FSX 21/05/2018 to 06/09/2018

LTP DIAGRAM AMENDED 1 RFA:- LDC REQUEST

Act:	Location:	Arr:	Dep:	Remarks:
PAO	Cardiff C		15.56	2M42 66789
(15.26 from BARRYIS to MERTHYR)				
	Cdf Q.St	15.59		

153
REL CV 446 at 16.09
on arrl 16.06 Cardiff Bay

	Cdf Q.St	16.09	16.12	2B96 2090 06277
	Cdf Bay	16.15		
	Cdf Bay		16.18	2S96 2051 43369
	Cdf Q.St	16.21		
	Cdf Q.St		16.24	2B98 2090 56291
	Cdf Bay	16.27		
	Cdf Bay		16.30	2S98 2051 03394
	Cdf Q.St	16.33		
	Cdf Q.St		16.36	2B00 2090 35308
	Cdf Bay	16.39		
	Cdf Bay		16.42	2S00 2051 72400
	Cdf Q.St	16.45		
	Cdf Q.St		16.48	2B02 2090 85322
	Cdf Bay	16.51		
	Cdf Bay		16.54	2S02 2051 32425
	Cdf Q.St	16.57		

RELD by CV 446 at 16.57

MRB Cdf Q.St 17.35 18.05

153
REL CV 446 at 18.09
on arrl 18.06 Cardiff Bay

	Cdf Q.St	18.09	18.12	2B16 2090 75464
	Cdf Bay	18.15		
	Cdf Bay		18.18	2S16 2051 22567
	Cdf Q.St	18.21		
	Cdf Q.St		18.24	2B18 2090 35489
	Cdf Bay	18.27		
	Cdf Bay		18.30	2S18 2051 72581
	Cdf Q.St	18.33		
	Cdf Q.St		18.36	2B20 2090 85503
	Cdf Bay	18.39		
	Cdf Bay		18.42	2S20 2051 32606
	Cdf Q.St	18.45		
	Cdf Q.St		18.48	2B22 2090 45528
	Cdf Bay	18.51		
	Cdf Bay		18.54	2S22 2051 82620
	Cdf Q.St	18.57		
	Cdf Q.St		19.00	2B24 2090 05553
	Cdf Bay	19.03		
	Cdf Bay		19.06	2S24 2051 42645
	Cdf Q.St	19.09		
	Cdf Q.St		19.12	2B26 2090 55567
	Cdf Bay	19.15		
	Cdf Bay		19.18	2S26 2051 02670
	Cdf Q.St	19.21		

RELD by TT 506 at 19.21
for 19.24 to Cardiff Bay

PAO	Cdf Q.St		19.26	2P91 65299
(18.48 from BARGOED to PENARTH)				
	Cardiff C	19.30		

143+143
REL CF 307 at 19.39
143
DP Front 2 cars

	Cardiff C	19D39	19D41	2A60 24907
(18.42 from BRGEND)				
FSX	Cdf Q.St	19.44	19.45	
	Cathays	19.47	19.48	
	Llandaf	19.52	19.52	
	Radyr	19.55	19.56	
	TaffsWell	19.59	20.00	
	Treffores	20.06	20.07	
	Pontyprid	20.10	20.12	
	Abercynon	20.21	20.21	
	Pnrhiwcbr	20.26	20.26	
	Mountain	20.30	20.33	
	Fernhill	20.35	20.35	
	Cwmbach	20.39	20.39	
	Aberdare	20.43		
	Aberdare		20.54	2P97 35363
FSX	Cwmbach	20.57	20.57	
	Fernhill	21.00	21.00	
	Mountain	21.03	21.04	
	Pnrhiwcbr	21.07	21.07	
	Abercynon	21.13	21.15	
	Pontyprid	21.22	21.24	
	Treffores	21.27	21.27	
	TaffsWell	21.34	21.34	
	Radyr	21.38	21.38	
	Llandaf	21.41	21.41	
	Cathays	21.45	21.46	
	Cdf Q.St	21.49	21.51	
	Cardiff C	21.54	21.56	2P97 35363
FSX	Grangtown	22.00	22.00	
	Dingle Rd	22.04	22.04	
	Penarth	22.07		
	Penarth		22.20	2F34 2202 89666
FSX	Dingle Rd	22.22	22.22	
	Grangtown	22.26	22.26	
	Cardiff C	22.31		

RELD by CF 315 at 22.31

SRB Cardiff C 22.36 22.51

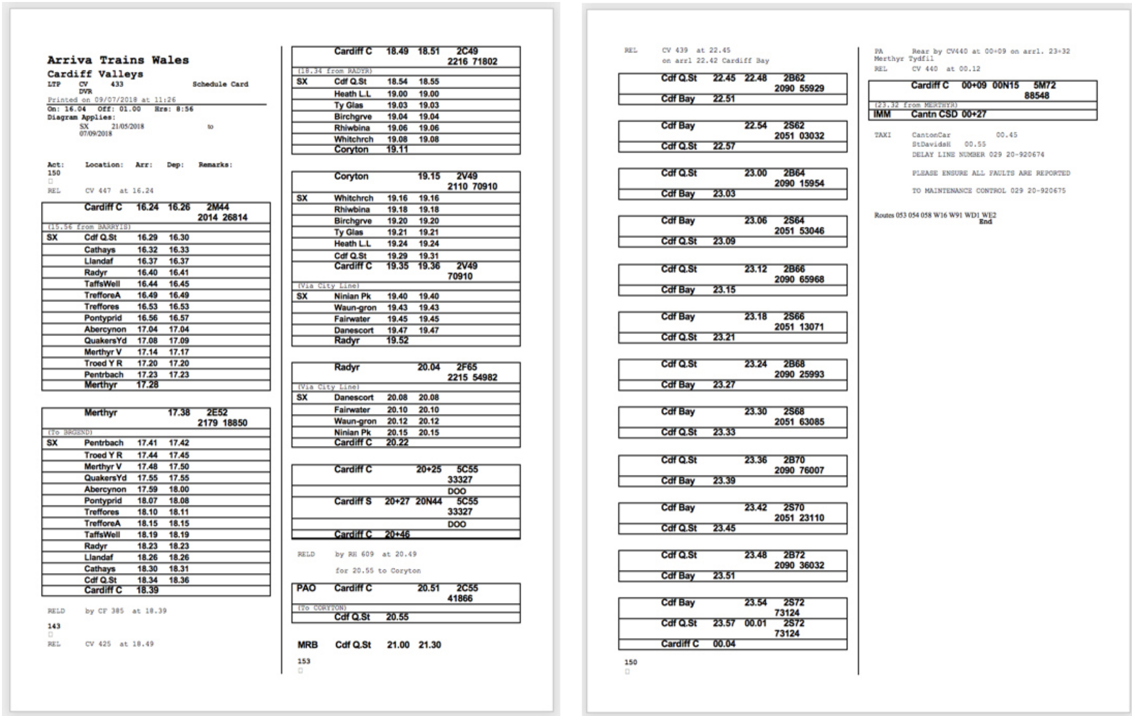
150
REL CV 436 at 23.01

	Cardiff C	23.01	23.12	2P99 85377
(21.54 from ABDARE)				
FSX	Grangtown	23.16	23.16	
	Dingle Rd	23.20	23.20	
	Penarth	23.23		
	Penarth		23.26	2F47 2202 54801
FSX	Dingle Rd	23.28	23.28	
	Grangtown	23.32	23.32	
	Cardiff C	23.39		

150
AP to unit in platform
RELD by CF 385 at 23.42
for ED to Canton DS
DELAY LINE NUMBER 029 20-920674
PLEASE ENSURE ALL FAULTS ARE REPORTED
TO MAINTENANCE CONTROL 029 20-920675

Routes 052 056 WE2
End

Figure 47: Diagram number: 448



Arriva Trains Wales
Cardiff Valleys
LTP CV 433 DVR Schedule Card
Printed on 09/07/2018 at 11:26
On: 16.04 Off: 01.00 Hrs: 8:56
Diagram Applies:
SX 21/05/2018 to 07/09/2018

Act: Location: Arr: Dep: Remarks:
150
REL CV 447 at 16.24

	Cardiff C	16.24	16.26	2M44 2014 26814
(15.56 from BARRYIS)				
SX	Cdf Q.St	16.29	16.30	
	Cathays	16.32	16.33	
	Llandaf	16.37	16.37	
	Radyr	16.40	16.41	
	TaffsWell	16.44	16.45	
	TrefforeA	16.49	16.49	
	Treffores	16.53	16.53	
	Pontyprid	16.56	16.57	
	Abercynon	17.04	17.04	
	QuakersYd	17.08	17.09	
	Merthyr V	17.14	17.17	
	Troed Y R	17.20	17.20	
	Pentrbach	17.23	17.23	
	Merthyr	17.28		

	Merthyr		17.38	2E52 2179 18850
(To BRGEND)				
SX	Pentrbach	17.41	17.42	
	Troed Y R	17.44	17.45	
	Merthyr V	17.48	17.50	
	QuakersYd	17.55	17.55	
	Abercynon	17.59	18.00	
	Pontyprid	18.07	18.08	
	Treffores	18.10	18.11	
	TrefforeA	18.15	18.15	
	TaffsWell	18.19	18.19	
	Radyr	18.23	18.23	
	Llandaf	18.26	18.26	
	Cathays	18.30	18.31	
	Cdf Q.St	18.34	18.36	
	Cardiff C	18.39		

RELD by CF 385 at 18.39
143
REL CV 425 at 18.49

	Cardiff C	18.49	18.51	2C49 2216 71802
(18.34 from RADYR)				
SX	Cdf Q.St	18.54	18.55	
	Heath L.L	19.00	19.00	
	Ty Glas	19.03	19.03	
	Birchgrve	19.04	19.04	
	Rhiwbina	19.06	19.06	
	Whitchrch	19.08	19.08	
	Coryton	19.11		

	Coryton		19.15	2V49 2110 70910
SX	Whitchrch	19.16	19.16	
	Rhiwbina	19.18	19.18	
	Birchgrve	19.20	19.20	
	Ty Glas	19.21	19.21	
	Heath L.L	19.24	19.24	
	Cdf Q.St	19.29	19.31	
	Cardiff C	19.35	19.36	2V49 70910
(Via City Line)				
SX	Ninian Pk	19.40	19.40	
	Waun-gron	19.43	19.43	
	Fairwater	19.45	19.45	
	Danescort	19.47	19.47	
	Radyr	19.52		

	Radyr		20.04	2F65 2215 54982
(Via City Line)				
SX	Danescort	20.08	20.08	
	Fairwater	20.10	20.10	
	Waun-gron	20.12	20.12	
	Ninian Pk	20.15	20.15	
	Cardiff C	20.22		

	Cardiff C		20+25	5C55 33327
				DOO
	Cardiff S	20+27	20N44	5C55 33327
				DOO
	Cardiff C	20+46		

RELD by RH 609 at 20.49
for 20.55 to Coryton

PAO	Cardiff C		20.51	2C55 41866
(To CORYTON)				
	Cdf Q.St	20.55		

MRB Cdf Q.St 21.00 21.30
153

REL CV 439 at 22.45
on arrl 22.42 Cardiff Bay

Cdf Q.St	22.45	22.48	2B62 2090 55929
Cdf Bay	22.51		

Cdf Bay		22.54	2S62 2051 03032
Cdf Q.St	22.57		

Cdf Q.St		23.00	2B64 2090 15954
Cdf Bay	23.03		

Cdf Bay		23.06	2S64 2051 53046
Cdf Q.St	23.09		

Cdf Q.St		23.12	2B66 2090 65968
Cdf Bay	23.15		

Cdf Bay		23.18	2S66 2051 13071
Cdf Q.St	23.21		

Cdf Q.St		23.24	2B68 2090 25993
Cdf Bay	23.27		

Cdf Bay		23.30	2S68 2051 63085
Cdf Q.St	23.33		

Cdf Q.St		23.36	2B70 2090 76007
Cdf Bay	23.39		

Cdf Bay		23.42	2S70 2051 23110
Cdf Q.St	23.45		

Cdf Q.St		23.48	2B72 2090 36032
Cdf Bay	23.51		

Cdf Bay		23.54	2S72 73124
Cdf Q.St	23.57	00.01	2S72 73124
Cardiff C	00.04		

150

PA Rear by CV440 at 00+09 on arrl. 23+32 Merthyr Tydfil
REL CV 440 at 00.12

	Cardiff C	00+09	00N15	5M72 88548
(23.32 from MERTHYR)				
IMM	Cantn CSD	00+27		

TAXI CantonCar 00.45
StDavidsH 00.55
DELAY LINE NUMBER 029 20-920674

PLEASE ENSURE ALL FAULTS ARE REPORTED

TO MAINTENANCE CONTROL 029 20-920675

Routes 053 054 058 W16 W91 WD1 WE2
End

Figure 48: Diagram number: 433

Observation for all four train driving instructors revealed and verified that beyond the initial setup procedure of keying-in the unit (i.e., mobilising the train and placing it on neutral), train drivers seemed to adopt the same preferred method – OTMR, GSM-R and DAS. This was not too surprising when considering that the OTMR system is the furthest away from the driver's seat located behind the passenger's seat, while the GSM-R and DAS are directly in front of the train driver (see Figure 49). However, it is important to highlight that login into the GSM-R is high priority, followed by the OTMR and only if time permits, as well as it being safe to do so, the DAS. Below is a brief description of the function of these three systems; OTMR, GSM-R and DAS in chronological order in which they were independently setup.

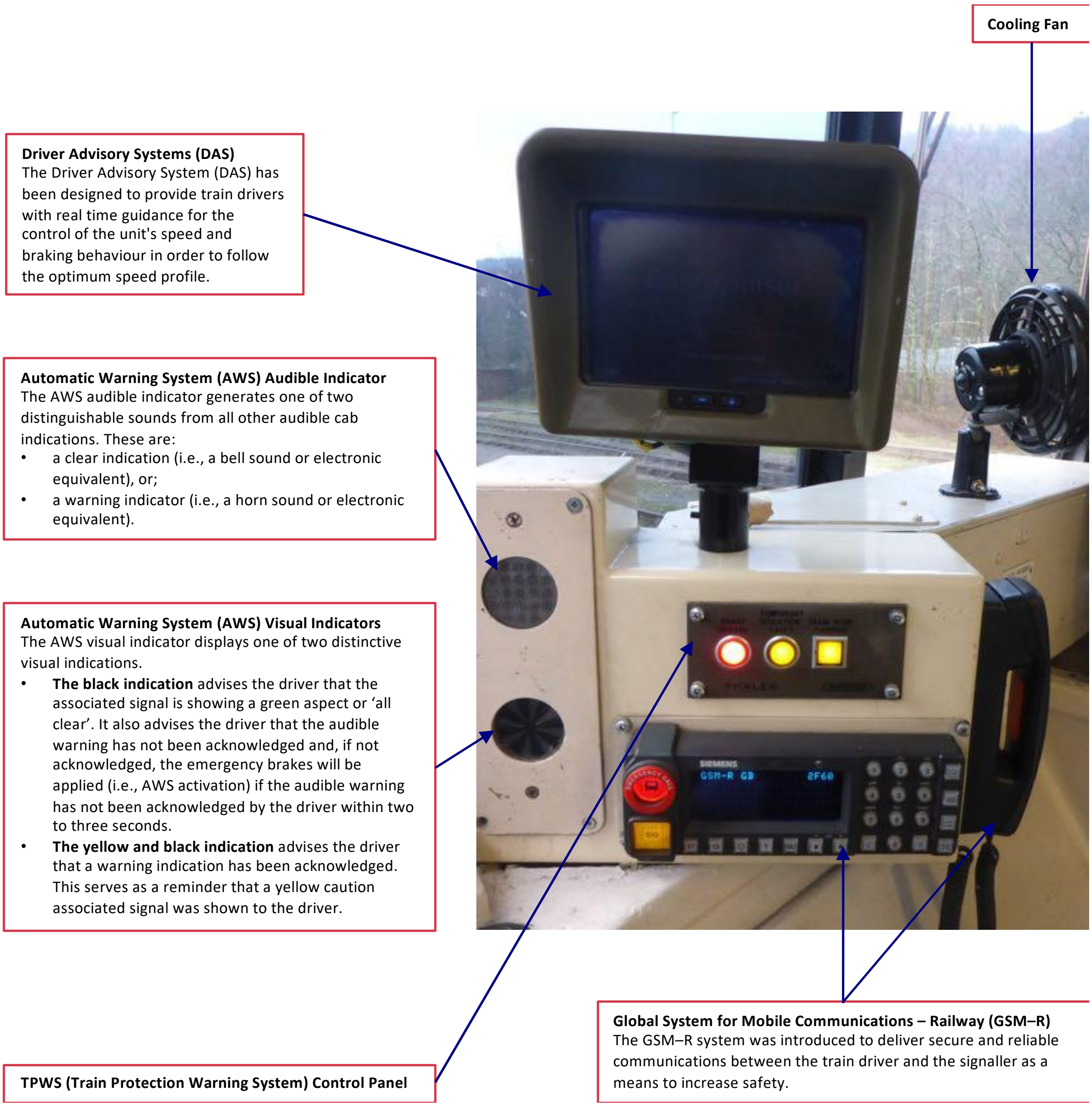


Figure 49: Global System for Mobile Communications – Railway (GSM-R), and Driver Advisory Systems (DAS)

4.8.2.1: On-Train Monitoring Recorder (OTMR)

The On-Train Monitoring Recorder (OTMR; see Figure 50, for a visual representation) provides a systematic safety monitoring system as a means of preventing incidents and accidents in Great Britain's mainline railway system (RSSB, 2014b). Observations revealed that the train driver instructors had to enter their driver's ID as well as their corresponding diagram section head code (see Section 4.7.2) into OTMR. Therefore, since a driver's ID and head code are entered into the OTMR, the ability to identify the driver, the unit and the infrastructure performance in the period leading up to and when possible directly after an incident or accident can be extracted to support the investigating team to better understand the circumstances leading up to the incident or accident – whether train driver related (i.e., human error) or unit equipment failure (RSSB, 2014b). As a result, the OTMR provides Arriva Train Wales (ATW) and other key stakeholders e.g., the Rail Safety and Standards Board (RSSB), the Office or Rail and Road (ORR), the Rail Accident Investigation Branch (RAIB), and Network Rail, with a complete digital record of each state change of all the monitored interactions within the unit signals e.g., Train Protection Warning System (TPWS), Automatic Warning System (AWS), Driver's Safety Device (DSD) pedal unit speed, and braking severity, depending on the severity of the accident. In addition, the OTMR recorded data can also be used to further support the case to review and potentially revise the current standard protocols and procedures in order to actively promote continuous development for change as a means to prevent future incidents or accidents from recurring (RSSB, 2014b). Therefore, the OTMR is the equivalent of the 'black box' that is installed on all aircrafts. Moreover, the OTMR employs proven techniques to provide protection of the recorded train data during accident conditions, which ultimately provides a secure data record for accident investigations (Arrowvale, 2018). However, the OTMR is more than a secure data recorder for the rolling stock fleet. The OTMR is also used by ATW to evaluate train drivers' operational performance and competency levels, whereby a train driver's manager can sit down with them and discuss their driving style to proactively reduce unforeseeable potential safety incidents as well as improve overall operational performance.

Figure 50: On-Train Monitoring Recorder (OTMR)

4.8.2.2: Global System for Mobile Communications – Railway (GSM–R)

In Great Britain's mainline railway system, the standards and rules state that train drivers and signallers must be able to communicate all at times (RSSB, 2018b). As a result, the Global System for Mobile Communications – Railway (GSM–R) system (see Figure 51, for a visual representation) was introduced to deliver secure and reliable communications between the train driver and the signaller as a means to increase safety (Network Rail, 2018; RSSB, 2017a) and has been adopted across five continents as the standard communication protocol (Chetty, Chen, & Woodbridge, 2016). The GSM–R is a constituent part of the ERTMS (European Rail Traffic Management System; Chetty *et al.*, 2016), which uses modern digital technology as part of the emerging 'digital railway' (RSSB, 2018c). In addition, the GSM–R has an alarm function that has been integrated into the Driver's Safety Device (DSD) pedal (AKA dead man's switch), which will automatically alert the signaller should the driver become incapacitated (Network Rail, 2012). It is important to point out that in-depth procedures have been developed and refined at Arriva Train Wales (ATW) to streamline and standardise communication between the train driver and the signaller (see RSSB, 2016a, for review) to further enhance safety as well as potentially reduce network disruptions, whether the GSM–R is utilised for normal point-to-point call, urgent point-to-point call, or railway emergency group call (REC; Network Rail, 2013; see Figure 51). Observations revealed that the train driver instructors had to enter their driver's ID as well as their corresponding diagram section head code (see Section 4.7.2) into GSM–R.

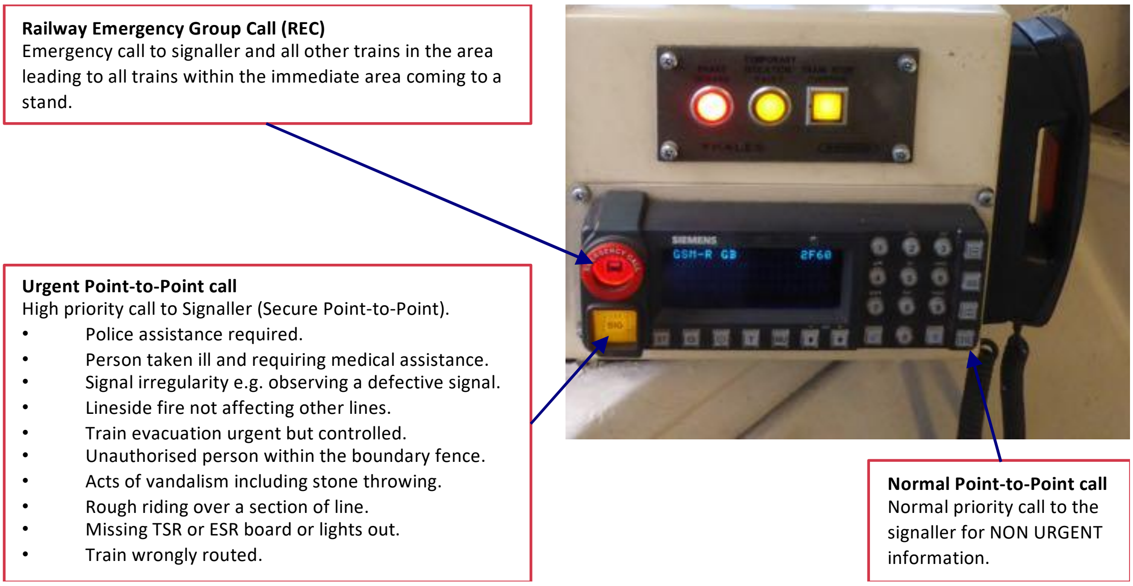


Figure 51: Global System for Mobile Communications – Railway (GSM–R)

4.8.2.3: Driver Advisory Systems (DAS)

The Driver Advisory System (DAS; see Figure 52, for a visual representation) has been designed to provide train drivers with real time guidance for the control of the unit's speed and braking behaviour in order to follow the optimum speed profile (Mitchell, 2018). At the point in time the observation took place, the DAS had only just begun to be rolled out, with financial incentives being provided to train drivers if they opted to interact and utilise the DAS unit under advisement. Therefore, the effectiveness of DAS as an advisory information system as a means to reduce energy wastage and increase network capacity without reducing the service quality or performance will require further research by Arriva Trains Wales (ATW). Observations revealed that the train driver instructors had to enter their driver's ID as well as their corresponding diagram section head code (see Section 4.7.2) into DAS.

Figure 52: Driver Advisory Systems (DAS)

4.8.3: Changing ends

On a diagram scheduling sheet (see Section 4.7.2) train drivers will always have to change ends at either the final station platform (e.g., Cardiff Central) or terminal station platform with a buffer stop (e.g., Swansea or Cardiff Bay). There are preparation and moving procedures in place that must be carried out at either the final station platform or terminal station platform with a buffer stop that train drivers must follow (See RSSB, 2017b; RSSB 2018d, for reviews), such as; platform speed restrictions, stopping distance from the signal or buffer stop as well as keying out from the unit end. Observation revealed that beyond the keying out from the unit and logging out of the Global System for Mobile Communications – Railway (GSM–R) system (see Section 4.8.2.2) and the Driver Advisory System (DAS; see Section 4.8.2.3) train driver instructors also changed the destination blind to reflect the final destination of the new section route as outlined on the diagram as well as changing the headlights to reflect the tail end (i.e., red) of the unit before finally stepping out of the cabin environment to visually inspect and verify the changes. The observer accompanied the train driver instructor throughout this process, which always included walking across the length of the platform in front of the train until a visual inspect and verification could be carried out in order to ensure that the tail end headlights were operational and indeed the tail end headlights. Furthermore, the train driver verified and inspected that the final destination was also displaying the new section route as outlined on the diagram. Upon completion, the external visual inspection of unit was carried out, simply to make sure there were no issues or damages that could potentially impact safety or operational function of the unit. At the other end of the unit, the reverse procedure was carried out. This also provided an opportunity for the train driver and guard to exchange information e.g., any operational issues or the guard letting the train driver know (s)he had changed the headlights and destination blind to reflect the new section route as outlined on the diagram. If this was done, then walking to the now front end of the unit to visually verify and inspect the headlights and destination blind was all that was needed. Otherwise, the train driver would need to go inside the cabin environment, change the headlights to reflect the front end and also change the destination blind to reflect the new section route as outlined on the diagram. This procedure was carried out every time the train driver had to change ends.

During observations, some diagrams from the Cardiff Valley line that had the very short journey for example from Cardiff Queen Street – Cardiff Bay return route required 20+ the procedures for changing ends (see Section 4.8.2 for examples from diagram number: 446 and diagram number: 433). Observations from these repetitive and intensive Cardiff Queen Street – Cardiff Bay diagrams revealed that having a preferred order in place to make sure that train drivers do not miss anything out has once again demonstrated and proven to be crucial in the events of showing signs of fatigue. Moreover, it was observed that it was also important to complete the changing ends procedure within the cab efficiently, especially on the Cardiff Queen Street – Cardiff Bay diagram section as during busy time periods exiting the unit can be somewhat challenging as passengers are trying to alight while others are trying to board the train. In some instances, those that are alighting can be quite rude, while those that are boarding tend to ask questions, such as; confirmation of destination, departure time or assistance on how to reach their final destination.

4.8.4: Preparing and moving a unit from Cardiff Shed Depot (CSD)

During prepping and moving a unit from Cardiff Shed Depot (CSD), beyond ensuring as an observer that the correct Personal Protective Equipment (PPE; ATW, 2017b), such as; steel toe cap shoes, hi vis vest and safety glasses were being used – there were an extensive list of operational safety protocols (see RSSB, 2017b; RSSB, 2018d, for reviews) that were demonstrated and explained in great detail. It is important to highlight that the observer did not hold any formal traction knowledge or relevant rolling stock (i.e., unit) technical skills to fully understand the operational safety protocols. Therefore, the observer will focus on outlining some of the issues faced while preparing and moving a unit from CSD that could be a contributing factor towards fatigue rather than the various steps required in prepping and moving a unit from CSD.

One of the biggest responsibilities of a train driver is to ensure that the unit that they are prepping and moving for the purpose of transporting passengers appears safe to travel, as well as the unit passing the various traction external inspections e.g.; fuel levels and air pressure, and internal safety inspections e.g., carriage emergency lighting system and fire detection system, including a complete sweep of the cab, e.g., Train Protection Warning System (TPWS), GSM–R, parking brake, Driver's Safety Device (DSD), Driver Reminder Appliance, horn and wiper isolator switches, vigilance, static brake test, static power test, emergency equipment (flags detonators and track circuit clips), public address, sand, traction interlock, emergency brake supply, front end and tail end headlights and internal lights, which should all be functioning within normal operating parameters. For example, train drivers must not allow a train to start a journey at any time during its planned working with a defective GSM–R in any cab, which is a major rule book requirement (RSSB, 2015f).

Train drivers are allocated 15 minutes to prepare to move a unit. During one observation, the unit the observer's train driver instructor was allocated to prepare was displaying the fire alarm warning system had an issue. Conversation between the observer's train driving instructor and the movement controller at CSD revealed that despite safety being portraited as the most significant factor within Arriva Trains Wales (ATW), the movement controller asked the observer's train driving instructor if it was ok for the unit to be taken out with the warning fire system light on. However, the observer's train driving instructor refused, and the movement controller allocated alternative unit, which required uncoupling from an original four carriage traction. Therefore, the unit was running far later than was originally allocated within the diagram. The original 10 minutes was absorbed by preparing the first unit, another 5 minutes was spent talking to the movement controller, and an additional 5 minutes uncoupling the new unit as well as having to restart all the procedures required for preparing a unit as stated above. As a result, the overseer's train driving instructor was running 15 minutes later than originally scheduled to the unit to be moved from SCD, which resulted in the signaller calling the observer's train driving instructor through the GSM–R asking why the unit was late. This whole process was quite stressful and set the tone for the whole day, which brings in rumination into the cab environment that could potentially increase safety critical incidents.

4.9: On the move

Rather than focusing on the standards and rules of the mainline railway system, the Train Operating Company's (TOC) unique protocols and procedures, traction knowledge, route learning/knowledge, etc., this section will discuss the various observations that were experienced by the observer that could be a contributing factor towards fatigue while the unit was on the move (see Figure 53, for a visual representation). Extensive in-cab observations by the observer (i.e., ~120 hours) revealed several reoccurring themes/issues that could be a contributing factor towards fatigue. These themes and issues were classified as; in-cab noises, in-cab temperature fluctuations, partial unit improvements, and in-cab working conditions. Below is a breakdown of the observer's experiences of each of these themes and issues.

Figure 53: Observer carrying out in-cab observations

4.9.1: Cab noises

While the unit was on the move, the observer had an extensive opportunity to experience first-hand some of the factors that could contribute towards fatigue. There were various noises within the cab environment that were quite overwhelmingly loud. Observations revealed that noise exposure varied depending on various environmental factors, such as traction (e.g., unit class type, maintenance, age, etc.), track condition, weather, and passenger behaviour (i.e., intoxicated passengers e.g., Friday and Saturday nights, or special events e.g., sport events and concerts) to name but a few. When experiencing the ill effects of fatigue this noise exposures adversely increase the likelihood of a train driver being involved in a safety incident whether due to distractions or an increase of experiencing noise sensitivity when irritated due to fatigue. Below are some of the observer's experiences relating to cab noises that were quite loud and distracting as well as drawing focus away from the required operational tasks.

4.9.1.1: Faulty window and door seals

There were issues in various units with faulty windows seals (e.g., Class 140s and Class 150s) as well as doors that did not close properly on some units (e.g., Class 150s). When the unit travelled at speeds greater than 45 miles per hour (mph) with the windows closed, the observer experienced continuous high pitch sounds inside the cab, changing in frequency and intensity depending on speed variation. Unfortunately, there was nothing that could be done to reduce the high pitch sounds. As an observer, time and cognitive resources were allocated in order to come up with an effective temporary solution to fix the seals and at least attempt to reduce the severity of the high pitch sounds. However, in doing so this resulted in a slight dip on the observer's situation awareness, whereby the observer had to confirm with the train driving instructor whether the train was safe as the observer was unsure whether the unit was still running on caution (i.e., on yellows). In such circumstance, the observer's train driving instructor explained the importance of the Automatic Warning System (AWS) indicator (See Section 2.5.2.2) as a means of double checking if the unit is running on caution.

4.9.1.2: Loud cooling fan

The observer noted that the cooling fan (see Figure 49) was very loud when switched on as well as not fit for purpose. On closer inspection, the observer was able to interact and experience first-hand the cooling fan's inability to generate sufficient power flow for the observer to even truly feel the air flow difference while sitting on the driver's seat, let along dropping the cab's working environmental temperature. The observer identified that the cooling fan was only able to gently circulate the warm air inside the cab. Therefore, the observer concluded that the trade-off between the loud noises that were generated by the cooling fan and the potential poor benefits of cooling down the train driver or cab's working environmental temperature was inadequate enough to merit the operational use of the cooling fan as it introduced further unwanted and potentially distracting noises into the cab's working environment that could contribute towards fatigue.

4.9.1.3: Other loud noises

The noise coming from the heater seems to have only two operational setting configurations – loud and louder, with no option to regulate a specific desired temperature (see Section 4.9.2). In addition, the air hydraulic windshield wipers aggressively banged side-to-side while in use, which was quite frustrating as well as annoying. Furthermore, the windshield wipers seemed to make a hissing sound in some units and in other units there was the common squeaking noise after three – five wipes when the screen was dry. There was no option available for the windshield wipers to function on intervals periods. Therefore, it was a constant process of switch-on and switch-off the windshield wipers, which impaired the observer's concentration and cognitive performance. These were a common occurrence for the observer across all the units observed (i.e., Class 142s, Class 143s, and Class 150s).

4.9.2: Cab temperature

There was no control in order to be able to regulate the temperature inside the cab. The observer found that the cab environment was quite a hostile environment to work inside. Based on the observer's limited number of diagram exposure, it was found that diagrams did not leave sufficient time beyond setting up the On-Train Monitoring Recorder (OTMR), the Global System for Mobile Communications – Railway (GSM-R; see Section: 4.8.2.2), and the Driver Advisory Systems (DAS; see Section: 4.8.2.3) before the unit was scheduled to leave. Therefore, all other subsequent cab environment adjustments, such as temperature control and seating alterations for ergonomic comfort, were an afterthought once the unit was on the move.

During very cold periods, especially during the winter nights, the observer felt as if he was inside an ice box. In contrast, during warm days the observer felt as if he was inside a greenhouse, despite the outside temperature being relatively mild. Observations throughout the four week period identified that none of the units observed had air conditioning. In addition, during observations, it became quickly apparent that it was vital for train drivers to have in their person a large bottle of water as well as a large thermal flask, whether filled with hot water, tea, or coffee. The observer realised early on during the first diagram that both were essential in mitigating the ill effect of the various unit's cab environmental temperature inconsistencies. For example, it was found that when the cab's starting internal temperature was below the optimal 16 degrees Celsius guideline under the Workplace (Health, Safety and Welfare) Regulations 1992, Approved Code of Practice and guidance, i.e., preparing and moving a unit from Cardiff Shed Depot (CSD; see Section 4.8.4), having the available option of a hot beverage, whether to hold or drink was a significant benefit for the observer to maintain focus when aiming to remain hydrated but the option of drinking cold water inside a cold cab environment was not appealing. Therefore, in order to reduce the potential impact of dehydration, having both cold and hot water was crucial.

4.9.3: Cab environment

With an ageing rolling stock fleet, there were clear signs of wear and tear across all units observed. Below are some of the observed wear and tear and cab environment issues that the observer felt that, when feeling the ill effects of fatigued, could be a contributing factor towards safety incidents.

4.9.3.1: Partial unit improvements

There was clear evidence of improvements having been integrated into various units. However, these improvements were clearly not even implemented across the same rolling stock class, let alone the whole rolling stock fleet. For example, it was noted by the observer that in some units the Automatic Warning System (AWS; see Section 2.5.2.2) acknowledgement pushbutton was either a silver pushbutton that is difficult to see in reduced lighting (i.e., in darkness) as well as having a poor ergonomic design, while other units had the better ergonomic AWS acknowledgement pushbutton design, which is more visible in reduced lighting due to it being larger in size as well as being bright yellow. It is also important to point out that in some units, the AWS acknowledgement pushbutton was a combination of both the silver and yellow AWS acknowledgement pushbutton design at opposite end of the same unit (see Figure 54).

Figure 54: Class 143 cab with different Automatic Warning System (AWS) acknowledgement pushbutton on opposite ends of the same unit

Observations also revealed that when changing from one unit to another, locating the AWS acknowledgement pushbutton required conscious awareness, which was not always possible as setting-up the cab always took precedence (see Section 4.8.2) and in some instances when the unit arrived behind the scheduled timetable departure, the observer noticed that the train driver quickly set-up before immediately being given the two-buzzer signal by the guard, which indicates their duties are complete and the train was 'ready to start'. Therefore, the observer felt that the location of the AWS acknowledgement button should be ergonomically integrated on the dashboards in the same location, which could potentially eliminate the AWS slow to cancel safety incidents. Moreover, the observer noticed that in the units that still had the old AWS acknowledgement silver pushbutton, some of these were so worn down that they had sunken further into the dashboard (see Figure 55). Furthermore, in some units the AWS acknowledgement silver pushbutton was flush with the dashboard or in some instances even below the dashboard (see Figure 56).

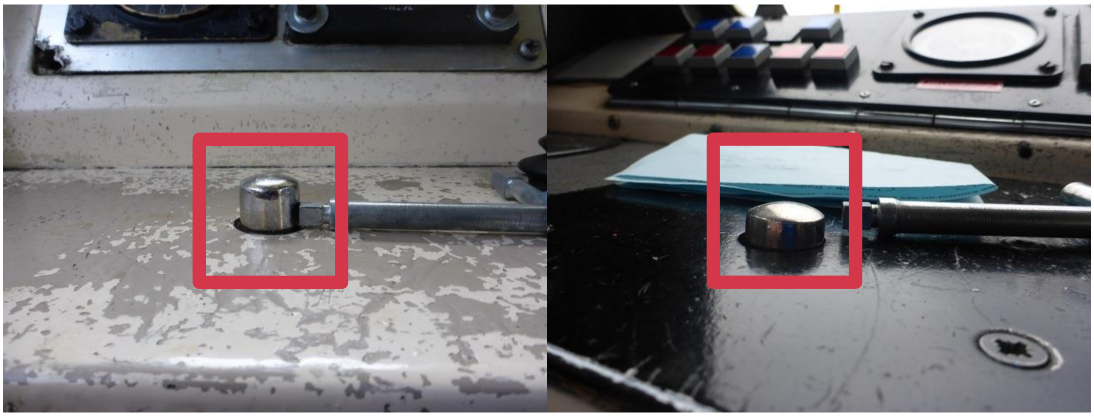


Figure 55: Automatic Warning System (AWS) acknowledgement silver pushbutton position variation

Figure 56: Automatic Warning System (AWS) acknowledgement silver pushbutton below the dashboard

4.9.3.2: Door that separates the cab from the saloon

In multiple occasions the observer experienced the door that is designed to separate the cab from the front saloon in the direction of travel would open unexpectedly whether on certain sections of the route (e.g., sharp bends) or either when applying speed or even braking. Closer inspections by the observer revealed the true state of the door lock mechanism on some of the units, which were clearly not fit for purpose (see Figure 57 and Figure 58). It would be reasonable for the observer to presume that entry to the cab should be restricted as well as providing an adequate level of security for the train driver. According to RSSB (2018e), train drivers must prevent unauthorised entry to the cab if possible, by locking the doors to the cab as well as other driving cab doors i.e., the tail end cab when changing ends (see Section 4.8.3). However, these door lock mechanisms are ineffective, which leaves train drivers vulnerable to passengers from the saloon who choose to enter the cab. This potential unauthorised access creates an unnecessary safety risk for both the train driver and passengers from individual with malignant intent to exploit limited security measures.

Figure 57: Example 1: Cab door

Figure 58: Example 2: Cab door

4.9.3.3: Other cab environment observations

There were other noteworthy issues inside the cab that the observer considered distracting and frustrating as well as drawing attention away from safely operating the unit. In some units the observer experienced light contamination from the destination blind (see Figure 59), which the observer felt was quite distracting during darkness. Furthermore, the condition of the cab clearly displayed ergonomic design failures to the extent where modifications were manually added to the control dashboard switches for further clarification and configuration of specific units (see Figure 60).

Figure 59: Light contamination from the destination blind

Figure 60: Cab light, destination light and heater setting configuration display

4.10: Discussion

The aim of this study was to identify some of the external environmental factors that could contribute towards safety incidents when fatigued through ethnographic research based on extensive in-cab observations. Cab observations revealed that noise (whether internal or external to the cab), cab temperature, and cab working conditions were major observational concerns that could contribute towards safety incidents when fatigued. According to Lipscomb and Roettger (1976), noise can cause annoyance and mental fatigue. There were various noises within the cab that were quite overwhelming for the observer. Cab observations identified that the various noise exposures were directly linked to the numerous environmental factors. These numerous noises resulted in the observer experiencing continuous distraction, irritation and lack of concentration. Melamed and Bruhis (1996) have identified that noise can cause physical fatigue and post-work irritability. However, despite the observer experiencing noise exposure habituation (Öhrström & Björkman, 1988; Öhrström, Björkman, & Rylander, 1990), when the observer began feeling fatigued, rather than noise exposure habituation occurring, noise sensitivity, irritation, distraction and anxiety increased, which significantly impaired the observer's subjective concentration levels, cognitive performance as well as the observer's judgement levels. Elmenhorst *et al.* (2014) identified that noise exposure significantly reduced cognitive performance. However, noise exposures have been found to be more long-term than simply reducing cognitive performance, whereby it can also affect job satisfaction (Sundstrom, Town, Rice, Osborn, & Brill, 1994). For example, Smith and Smith (2017b) found that at ATW, noise exposure had a negative effect on well-being. In addition, van Kamp and Davies (2013) state that individuals with a mental health condition experience higher noise sensitivity as well as reporting higher discomfort levels. Furthermore, Beutel *et al.* (2016) have argued that noise exposure is associated with a two-fold higher prevalence of anxiety and depression when compared to the general population.

Seider *et al.* (2016) have found that noise exposure on railway staff results in poor health outcomes. For example, Lie, Skogstad, Johnsen, Engdahl and Tambs (2016) identified that rail staff developed hearing loss as a result of noise exposure. However, Lie *et al.* (2013) have pointed out that Norwegian train drivers and conductors (i.e., guards) had normal hearing threshold levels comparable with those in non-exposed groups. Further conflicting evidence in the literature seems to be split regarding whether noise exposure in the rail industry results in hearing loss (Clark & Popelka, 1989; Henderson & Saunders, 1998; Kryter, 1999; Malleson, 1989). However, in Lie, Skogstad, Johnsen, Engdahl and Tambs' (2013) study audiogram data was obtained from the electronic medical records of the Occupational Health service together with information on age, gender and type of job but no information was collected from the Train Operating Company (TOC), such as; knowledge of route unit (i.e., traction), track rail (i.e., rail), cab environment, and age of the rolling stock fleet, which were clear issues and contributing factors towards fatigue within Arriva Trains Wales (ATW), based on observations. Therefore, it was observed that a large proportion of noises could be reduced or even eliminated from the cab through effective maintenance or introduction of new rolling stock. Lie *et al.* (2013) highlights that noise exposure levels are dependent on the quality and maintenance of both the traction and rail. Therefore, it seems that noise exposure when fatigued could adversely increase the likelihood of a train driver being involved in a safety critical incident. However, noise was not the only observed external environmental factor that could contribute towards safety incidents when fatigued. Cab temperature was also another major concern observed.

There were no internal controls installed inside the cab in order to be able to regulate the temperature. The observer found that the cab environment was quite a hostile environment to work in. Based on the observer's limited number of diagram observations (13 diagrams in total, see Table 4, Section 4.3.2), it was found that the diagram planning team at ATW did not leave sufficient time beyond setting-up the primary systems e.g., the OTMR, the GSM–R, and the DAS before the train was scheduled to leave. Therefore, all other subsequent cab environment adjustments, such as temperature control and seating alteration for ergonomic comfort, were an afterthought once the train was on the move. Research carried out by Hancock *et al*. (2007) and Pilcher *et al*. (2002) have highlighted that working in train cabs during hot and cold temperatures can result in increased risks and a reduction in human performance tasks such as; reduced concentration, tunnelling of vision, reduced vigilance, reduced work rate, and a slower performance on repetitive tasks. In addition, Qian *et al*. (2015) have identified that heat has a potential fatigue-enhancing effect when individuals are performing highly cognition-demanding attention task. Therefore, without the ability to effectively and consistently regulate the cab temperature, it quickly became apparent that effective coping strategies were implemented and managed. As a result, based on observations it was vital for train drivers to have in their person a large bottle of cold water as well as a large thermal flask (e.g., hot water, tea, or coffee). The observer realised during the first diagram that both cold and hot water were essential in mitigating the ill effect of the various unit's cab environmental temperature inconsistencies. Research has shown that mild hypohydration can cause symptoms of fatigue (Gisolfi & Copping, 1993; Nielsen *et al*., 1993; Nielsen, Strange, Christensen, Warberg, & Saltin, 1997), pain (Ogino, Kakeda, Nakamura, & Saito, 2014), and cognitive performance (Adan, 2013; Ganio *et al*., 2011) as well as a significant increase in minor driving errors during a prolonged, monotonous drive (Watson, Whale, Mears, Reyner, & Maughan, 2015). However, both noise exposure and cab temperature were one aspect from external environmental factors that could contribute towards safety incidents when fatigued. Observations also revealed that the cab environment was also another major concern.

With an ageing rolling stock fleet at ATW, there were clear signs of wear and tear across all units observed. Whether these issues were due to extensive handling exposure e.g., dashboards, switches, window and door seals or simply natural ageing e.g., rust or broken doors – the cab environment was deemed by the observer to be quite run down. However, despite the ageing rolling stock, ATW seemed to have taken an active effort to modernise safety systems whether for ergonomic improvements e.g., Automatic Warning System (AWS) and yellow AWS acknowledgement pushbutton (RSSB, 2015b) or efficiency e.g., integrating the DAS, which provides real time guidance for the control of the unit's speed and braking behaviour in order for the train driver to follow the optimum speed profile (Mitchell, 2018). Nevertheless, observations revealed that at ATW, the AWS acknowledgement pushbutton was not upgraded from the poor ergonomic design of the silver AWS acknowledgement pushbutton to the better ergonomic design of the yellow AWS acknowledgement pushbutton. More surprising for the observer was the fact that some units had both the silver and yellow ergonomic AWS acknowledgement pushbutton design at opposite ends of the same unit. Moreover, it was also observed that the location of the AWS acknowledgement pushbutton was in different locations across the rolling stock fleet (e.g., Class 142s, Class 143s, etc.), which is not too surprising but what was surprising was that in some units, the AWS acknowledgement pushbutton was in a different location within the same rolling stock class. Furthermore, in some units the AWS acknowledgement silver pushbutton was flush with the dashboard and in other units even below the dashboard. For these units, train driver’s reaction time could be significantly increased, especially since fatigue results in impaired alertness (Åkerstedt & Ficca, 1997; Dorrian, Lamond, & Dawson, 2000; Gillberg *et al.*, 1994; Sussman & Coplen, 2000). This could potentially result in an AWS slow to cancel safety incident since the AWS needs to be acknowledged by the train driver within two to three seconds (RSSB, 2015b). However, external environmental factors such as noise exposure, cab temperature, and work conditions were one aspect that could contribute towards safety incidents when fatigued. Observations also revealed that work pressure whether direct or indirect could also contribute towards safety incidents when fatigued.

According to Zoer, Ruitenburg, Botje, Frings-Dresen and Sluiter (2011), work pressure was the most significant risk factors of railway employees for expressing mental health complaints. For example, if the train was running more than three minutes late, control would contact the train driver via the Global System for Mobile Communications – Railways (GSM-R) to establish whether the delay was due to; the train driver, another division within ATW or external factor e.g., signaller, fleet, station staff, guard, Network Rail infrastructure faults e.g., signal box, cable theft, trespass, etc. or alternatively passenger related delays, e.g., overcrowding on commuter services (i.e., delays due to passengers alighting and boarding), wheelchair accessibility, anti-social behaviour, and on-board incident. Based on observations, it was felt that train drivers were working exhaustively hard to make up time where possible within the scope of the rules and regulations in order to ensure the train operated as close as possible within the established train operating company's (TOC) timetable. However, based on the researcher's unique experience, it was felt that the diagrams were designed to maximise performance based on ATW's timetable without acknowledging or tailoring and revising each route to reflect reoccurring delays such as passenger related delays, e.g., overcrowding on commuter services (i.e., delays due to passengers alighting and boarding), wheelchair accessibility, or anti-social behaviour. The observer felt that more could be done to actively tailor routes based on frequently recorded delays. For example, anti-social behaviour is more common on specific routes towards the evening as well as more prevenient on the weekend. Beyond increasing onboard security presence personnel, perhaps other methods could be integrated to complement current strategies such as the introduction of tailored deterrent onboard public voice announcements. However, alternative strategies could be costly to implement and maintain as well as redirecting valuable resources. Based on observations, perhaps a review could be commissioned to determine each route's reoccurring delays so that diagrams could be revised to factor in such delays going forward. Nevertheless, this ethnographic research was not without limitations.

One of the biggest limitations with ethnographic research is that the observer can become biased towards the direction of the research i.e., 'observer effects' (LeCompte & Goetz, 1982). However, Monahan and Fisher (2010) argues that ethnographic research lies in cultivating close ties with others in the field. Therefore, the researcher had the opportunity to observe train drivers as well as experience first-hand some of the challenges encountered that contribute towards safety incidents when fatigued from within the cab environment. This process provided a unique perspective as well as providing a far deeper understanding of the role and the various unforeseeable circumstances that the observer had not envisioned. As a result, the researcher was able to go beyond making assumptions in the comfort of a well-lit and temperature-controlled environment regarding what are some of the challenges that train drivers experience on a daily base that could contribute towards safety incidents when fatigued. However, it is also important to acknowledge another major limitation of conducting ethnographic research, which was information that was either true or false from the train driving instructor's perspective or as ambiguous and uncertain from one or both perspectives. According to van Maanen (1979), false and misleading information is valuable to the observer when it is recognised to be false. Therefore, it was important for the researcher to examine the ethnographic observations and when possible seek clarification from the train driving instructor at the earliest convenience e.g., during the next observation or when the train driving instructor is booking-on for their next shift beyond the period of the observations. Further research is now needed to determine whether other objective indicators of fatigue could be developed and validated to better support the Fatigue Risk Management System (FRMS) at Arriva Trains Wales as previous chapters have identified that the current Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator used by ATW may not be an effective predictor of train driver's fatigue as well as cab observations highlighting that there are several external environmental factors (e.g., noise exposure, inability to regulate cab temperature, work conditions and environment) that the current HSE's FRI calculator does not address or acknowledge that could contribute towards safety incidents when fatigued.

4.11: Chapter summary

The aim of this study was to identify some of the external environmental factors that could contribute towards safety incidents when fatigued through ethnographic research based on extensive in-cab observations. The observer was paired up with experienced train driving instructors who were highly competent in operating trains for the carriage of passengers as well as being able to carry out their duties while being accompanied by a third party inside the cab. The observer completed ~120 hours of consecutive in-cab observations. Through the pairing, the observer was able to observe, discuss and reflect as well as obtaining fundamental rules and regulations knowledge, traction knowledge, and route knowledge, which also included a comprehensive understanding of the practical train handling – all obtained within the cab. Through the process of ethnographic research, the observer was able to identify that noise exposure (whether internal or external), cab temperature, and cab working conditions were major observational concerns that could contribute towards safety incidents when fatigued. However, it was observed that a large proportion of the identified external environmental factors could be reduced or even eliminated through effective maintenance or through the introduction of new rolling stock. As a result, it is recommended that ATW considers that any new rolling stock added to the current fleet incorporates significant soundproofing in the cab as well as an effective air conditioning system that permits cab temperature control for train drivers. Moreover, it is recommended that ATW's diagram planning team identifies for each train journey the most frequently recorded delays e.g., overcrowding on commuter services (i.e., delays due to passengers alighting and boarding), wheelchair accessibility, anti-social behaviour, and on-board incident, etc. in order to be able to generate diagrams that realistically and proactively reflect operation as well as produce more accurate timetables that meets customer's transparency expectations. Further research is now needed to determine whether other objective indicators of fatigue could be developed and validated to better support the Fatigue Risk Management System (FRMS) at Arriva Trains Wales as previous chapters have identified that the current Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator used by ATW may not be an effective predictor of train driver's fatigue, as well as cab observations highlighting that there are several external environmental factors (e.g., noise exposure, inability to regulate cab temperature, work conditions and environment) that the current HSE's FRI calculator does not address or acknowledge that could contribute towards safety incidents when fatigued.

Chapter 5: Developing and validating an alternative online objective mobile indicator of fatigue

5.1: Overview of chapter

The previous chapter identified external environmental factors that could contribute towards safety incidents when fatigued through ethnographic research. Extensive cab observations revealed that; noise (whether internal or external to the cab), cab temperature, and cab working conditions were major concerns that could contribute towards safety incidents when fatigued. The aim of this chapter is to develop and validate a cognitive neurobehavioral performance measure for monitoring temporal dynamic changes in sustained attention that could be used to provide an objective indicator of fatigue in frontline safety critical workers such as train drivers, hospital staff, emergency services, law enforcers, etc. This is crucial as it was previously found that the current biomathematical model (BMM) of fatigue used at ATW i.e., the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator may not be an effective or accurate predictor of frontline safety critical worker's fatigue levels.

5.2: Introduction and rationale

In a controlled laboratory setting, the human Psychomotor Vigilance Task (PVT) (see Dinges, Orne, Whitehouse & Orne, 1987; Dinges & Powell 1985, for reviews) has become the widely accepted 'gold standard' tool for assessing the impact of sleep deprivation and fatigue on human cognitive neurobehavioral performance for monitoring temporal dynamic changes in attention (e.g., Belenky *et al*., 2003; Dinges *et al*., 1997; Jewett, Dijk, Kronauer & Dinges, 1999; Lamond *et al*., 2003). In retrospect, the PVT could be traced back from the early work in simple reaction time (SRT) studies that were carried out by Wilhelm Maximilian Wundt (1832 – 1920), which was then continued by James McKeen Cattell (1860 – 1944) (Davis, Roma, & Hienz, 2016) and then alternative early versions of a reaction time test, which used magnetic tape memory (Wilkinson, &

Houghton, (1975). It is important to note that the modern PVT has been refined several times over the years from its original development by Dinges and Powell (1985) (e.g., Basner & Dinges, 2011; Basner, Mollicone, & Dinges, 2011; van Dongen & Dinges, 2005), and has been shown to be sensitive to sleep deprivation, fatigue, drug use, and age. The PVT has been widely implemented using a handheld device (see Figure 61) known as the PVT-192 (Ambulatory Monitoring Inc., Ardsley, New York, USA), as well as being extensively validated by various researchers (Basner & Dinges, 2011; Dorrian, Roach, Fletcher & Dawson, 2007; Lamond, Dawson & Roach, 2005; Lamond *et al*., 2008; Loh, Lamond, Dorrian, Roach & Dawson, 2004; Roach, Dawson & Lamond, 2006).

Figure 61: PVT-192 Psychomotor Vigilance Task Monitor

According to Basner, Mcguire, Goel, Rao and Dinges (2015) and Dorrian, Rogers and Dinges (2005), the PVT-192 records participants sustained attention based on repeated reaction time (RT) trials to visual stimuli that occur at random inter-stimulus intervals (ISI) that are between 2–10 seconds, for a standard 10-minute period. In summary, the PVT-192 device operated by presenting participants with a stimulus that consisted of a four-digit millisecond counter that appears in a light-emitting diode (LED) dot-matrix display. The response consisted of a left or right button press, which depended on the configuration of the PVT-192 setup. The time difference between the stimulus presentation and the response constituted the participant's reaction time (RT). Each RT value was stored in the device and then uploaded to a personal computer, where the individual RTs are post processed with the REACT software (Ambulatory Monitoring Inc., Ardsley, New York, USA), or other commercially available software, into summary statistics, such as the mean RT or the mean number of lapses (RTs ≥500 milliseconds) per session (Basner & Dinges, 2011; Dinges & Kribbs, 1991; Dinges & Powell, 1985; Dorrian *et al.*, 2005; Warm, Parasuraman & Matthews, 2008). As an exemplar, in Roach *et al.*'s (2006) study, each participant performed either 5 minutes or 10 minutes RT sessions spaced at predetermined intervals (e.g., every 2 hours) for a prolonged duration (e.g., 28 hours), where each session consists of either 50 trials (equivalent to 5 minutes), or 100 trials (equivalent to 10 minutes). However, Khitrov *et al.* (2014) tested the average delay of the PVT-192 and found that their recorded delay was greater than what was stated by the PVT-192 manufacturer. Their recorded delay found that on average, it was 2.4 ms greater when compared to the manufacturer's reported delay of 1 ms. Nevertheless, it is important to highlight that Khitrov *et al.* (2014) did acknowledge the possibility that the difference found could have been due to the non-instantaneous nature of the light detection circuit, or the actual delay associated with the PVT-192, since their experimental design did not permit them to be able to distinguish between these possibilities.

Dinges and Powell (1985) have shown that the 10-min PVT is highly reliable. Roach *et al.* (2006) wanted to investigate whether 90 seconds could also be sufficiently sensitive enough to detect the effects of fatigue in comparison to their earlier research (see Loh *et al.*, 2004, for review), where it was possible to find significant fatigue-related impairment during the first 5-min of a 10-min PVT. In this study, the researchers compared participants' neurobehavioral performance using the PVT between three different time durations (90 seconds, 5-min, and 10-min) to identify whether a shorter PVT could also be sensitive enough to detect the effects of fatigue. They found that it was only possible to implement a 5-min PVT as a substitute of the 10-min PVT, and not a 90 seconds PVT, thus only further supporting their earlier research (i.e., Loh *et al*., 2004). However, it is important to note that analyses of their study were carried out using the mean RTs and not the mean speed responses (1/RTs). Basner and Dinges (2011) have identified that the mean RTs should not be the primary measure of alertness, and instead consider using the alternative primary measure of 1/RTs. Analyses of journal manuscripts reporting PVT results, published between 1986 – 2010 (n = 141) showed that there was great variability in the use of the PVT outcome metrics (see Basner & Dinges, 2011, for review), with the most frequently PVT outcome analyses; 60.7 per cent reporting mean number of lapses, 40.4 per cent reporting mean reaction time (RT), and 30.5 per cent reporting mean speed response (1/RT). It is important to highlight that 1/RTs are calculated using the following equation (Belenky *et al*., 2003):

5.2.1: Speed response (1/RT) equation:

$$\text{Speed Response } (1/\text{RT}) = \left(\frac{1}{\text{Reaction Time}}\right) \times 1000$$

In a later study, Basner, Mollicone and Dinges (2011) aimed to further shorten the 5-min PVT (i.e., Roach *et al*., 2006) by developing a modified 3-min version of the PVT (PVT-B). They found that their 3-min version could be a useful tool for assessing behavioural alertness in settings where the ‘gold standard’ 10-min PVT could be more difficult or impractical to implement due to the nature of the study or location. Therefore, further validation is required to determine whether both the 5-min PVT and PVT-B versions could indeed be sensitive enough to detect reduced levels of fatigue and on alternative mobile devices. However, it is important to point out that the various PVT versions were administered either on the PVT-192 or a personal computer.

According to Barnhoorn, Haasnoot, Bocanegra and van Steenbergen (2015), the ability to run online behavioural experiments that require precise recording of reaction times (RT) tend to be significantly more complex to achieve. Nevertheless, Crump, McDonnell and Gureckis (2013) and Reimers and Stewart (2016) were able to demonstrate and replicate RT experiments using JavaScript and HTML (hypertext markup language), as well as other studies demonstrating comparability between lab based and online based experiments (e.g., de Leeuw & Motz, 2016; Reimers & Stewart, 2007; Schubert, Murteira, Collins, & Lopes, 2013; Simcox & Fiez, 2014). At present, alternative online experiment solutions include; The Qualtrics Reaction Time Engine (QRTEngine) (Barnhoorn *et al.*, 2015), Training and Testing Tool (Tatool) (von Bastian, Locher, & Ruflin, 2013), WebExp (Keller, Gunasekharan, Mayo, & Corley, 2009), and ScriptingRT (Schubert *et al.*, 2013). These experiments were created in order to provide alternative online accurate timing experiments. Barnhoorn *et al.* (2015) have outlined that these libraries can be utilised to create RT experiments that have both cross-platform as well as cross-browser compatibility. However, with the exception of QRTEngine, such online experiments require incredible programming proficiency skills and software installation packages. Conversely, according to one of the developers of QRTEngine, E. Haasnoot (personal communication, February 26, 2016) outlined that:

> '...*Unfortunately, Qualtrics is pretty unresponsive to queries about 3rd-party JavaScript solutions. As far as I know, the QRTEngine works at the moment, but we have no idea when they will be introducing more breaking changes. Building your own task that doesn't necessarily rely on Qualtrics to work would be the safer bet...*'

In addition, due to unpredictable developments in the Qualtrics environment, which made it increasingly difficult for the team to provide a stable QRTEngine platform, in June

2016, the team decided to retire QRTEngine as well as their support (van Steenbergen, 2016).

Evans, Harborne and Smith (2019) developed an alternative online mobile version of the Psychomotor Vigilance Task (PVT) i.e., online mobile Psychomotor Vigilance Task (online m-PVT). The m-PVT was administered on two distinctive mobile devices – Apple's iPhone 6s Plus and Samsung Galaxy Tab 4 counterbalanced for both the morning and afternoon sessions i.e., time-of-day effect (see Smith, 1992, for review) and for the duration of 25 minutes i.e., time-on-task effect (see Langner & Eickhoff, 2013; Mackworth, 1948; Mackworth, 1968, for reviews). In their study it was found that Apple's iPhone 6s Plus generated reaction times (RTs) that were comparable to those found in the literature (e.g., Basner *et al*., 2011; Basner & Dinges, 2011; Dinges *et al*., 1987; Dinges & Powell, 1985; Dorrian *et al*., 2007; Evans *et al*., 2019; Khitrov *et al*., 2014; Lamond *et al*., 2005; Lamond *et al*., 2008; Loh *et al*., 2004; Roach *et al*., 2006), while the Samsung Galaxy Tab 4 generated significantly slower RTs than Apple's iPhone 6s Plus mobile device and thus slower than those comparable to the literature. However, it is important to note that based on Basner and Dinges (2011) recommendations of removing RTs < 100 ms (i.e., false start) and RTs ≥ 500 ms (i.e., number of lapses), Evans *et al.* (2019) identified that there was a significantly higher number of lapses for the Samsung Galaxy Tab 4 when compared to Apple's iPhone 6s Plus. This was not too surprising when considering that since participants mean RTs were slower for the Samsung Galaxy Tab 4 – it would be expected to find a significantly larger proportion of RTs that exceed Basner and Dinges (2011) RT cut-off (i.e., ≥ 500 ms) threshold. Moreover, using the time-on-task effect, Evans *et al.* (2019) revealed that Apple's iPhone 6s Plus based on RT analyses were sensitive enough at detecting levels of fatigue after 10 minutes on the task, while the Samsung Galaxy Tab 4 based on number of lapse analyses were sensitive enough at detecting levels of fatigue after only five minutes on the task.

Nevertheless, this study did not find significant diurnal variations between the morning and afternoon administration of the online m-PVT i.e., time-of-day effect, despite it having been identified that performance is subject to diurnal variations (Lenne, Triggs, & Redman, 1997). However, only 26 participants were recruited and split into two mobile devices groups. Therefore, perhaps 13 participants may not provide enough power to detect a time-of-day effect based on a 2 × 2 × 6 mixed-design analysis of variance (ANOVA) with 2 × mobile devices (Apple iPhone 6s Plus, or Samsung Galaxy Tab 4) as the between-subjects factor, and × 2 time of day (morning, or afternoon) × 6 time on task (1-minute; 5-minutes, 10-minutes, 15-minutes, 20 minutes, or 25-minutes). Nevertheless, it is important to acknowledge that Evans *et al.'s* (2019) study was not carried out to explore the time-of-day effect, but to determine the most appropriate mobile device to explore an alternative online mobile version of the 'gold standard' 10-minute Psychomotor Vigilance Task (10-min PVT) that could be used to provide an objective indicator of fatigue for frontline safety critical workers such as train drivers.

Carrier and Monk (2000) carried out a review of time of day on performance and identified that human performance efficiency changes as a function of time of day. The effect of time of day on performance have been studied extensively (see Folkard, 1983; Folkard & Monk, 1985, for reviews). For example, Patkai (1971) investigated diurnal variations in alertness, performance, and adrenaline excretion and found that adrenaline excretion was highest for morning workers and decreased gradually during the day, while evening workers showed nearly constant adrenaline excretion. In addition, it was also found that performance did not vary during the day for morning workers, while evening workers showed a steady improvement. Moreover, Lafrance and Dumont (2000) examined diurnal variation in temperature, mental and physical performance in university football students and found that self-reported alertness ratings increased, while self-reported fatigue ratings decreased throughout sessions that were spaced at predetermined intervals (e.g., every 4 hours) for a prolonged duration (e.g., 12 hours; 08:00 – 20:00). These findings further support the time-of-day effect. However, Smith (1992) found that performance was faster but less accurate in the early evening when compared to the early morning. Therefore, it has been found that performance on the PVT is sensitive to both time awake and circadian rhythms, which makes it ideal for providing an objective measure of cognitive function (e.g., fatigue, workload, alertness, etc.) (Gunzelmann, Moore, Gluck, van Dongen, & Dinges, 2011). As a result, the aim of this study is to replicate and validate using the time-of-day and time-on-task paradigm to determine whether an alternative online mobile version of the 'gold standard' 10-minute Psychomotor Vigilance Task (10-min PVT) i.e., online 10-minute mobile Psychomotor Vigilance Task (online 10-min m-PVT) could be used to provide an objective indicator of fatigue for frontline safety critical workers such as train drivers.

5.3: Methodology

5.3.1: Ethical approval

The study received ethics approval from Cardiff University's Ethics Committee (EC.16.02.09.4457A). The study conformed to the seventh amendment of the Declaration of Helsinki 1964 (World Medical Association, 2013) and was in accordance with the Data Protection Act 1998 and the General Data Protection Regulation (GDPR). All participants gave their Informed written consent (counterbalanced order: morning / afternoon, Appendix A; afternoon / morning, Appendix I) as well as electronic consent following the explanation of the nature of the study.

5.3.2: Participants

74 (9 male and 65 female) mean age 19.41 (*SD* = 1.61) participants were voluntarily recruited from Cardiff University via the Experimental Management System (EMS) to take part in the study. Participants were instructed not to consume caffeine (e.g., coffee,

energy drinks, tea, etc.) and alcohol during the 24 hours before the study. The study involved participants attending two counterbalanced sessions, a morning session (i.e., before 11:00) and an afternoon session (i.e., after 17:00), which were held on two consecutive days, in exchange for £10 or partial course credits. The study lasted 60 minutes in total for both sessions.

5.3.3: Statistical analyses

IBM's Statistical Package for the Social Sciences (SPSS) version 23 for Mac was used to analyse the data. A combination of various statistical procedures were carried out on the data; intraclass correlation coefficient (ICC) analysis, descriptive analyses, mixed-design analysis of variance (ANOVA) and a two-way analysis of variance (ANOVA) to further explore interactions. The level of $\alpha < .05$ was used for all statistical tests of this experiment.

5.3.4: Materials and apparatus

The mobile Psychomotor Vigilance Task (m-PVT) was presented to participants on an Apple iPhone 6s Plus running Apple's iOS version 9.3.1 (Apple Inc.). The m-PVT run in the following hardware configurations; system chip (Apple A9 APL1022), processor (Dual-core, 1840 MHz, Twister, 64-bit), graphics processor (PowerVR GT7600), and system memory (2048 MB RAM). The m-PVT was displayed on a 5.5-inch (diagonal) 1920 × 1080-pixel native resolution at 401 ppi Retina high definition display. The m-PVT was programmed using the client code HTML (hypertext markup language), and CSS (cascading style sheets) for the page visualisation and layout. JavaScript was also used to initiate the online m-PVT, which was run using the Dolphin Web Browser (MoboTap Inc.) on the Apple's iPhone 6s Plus (Dolphin Web Browser version 9.9.0, released August 2011). The rational for selecting the Dolphin Web Browser for this study was that it allowed the full screen feature to be enabled, while other more native internet browsers did not, such as Safari, Chrome, and Firefox to name but a few. The online Qualtrics Surveys (Qualtrics Labs, Inc. version 13.28.06) was also used to collect demographic information. In order to increase validity and standardisation, all instructions were administered to participants in written forms (see Appendix B, C, D, E, F, and G, for morning / afternoon order; and see Appendix J, K, L, M, N, and O, for afternoon / morning order). Participants were also verbally debriefed at the end of the study to explain the nature of the study in accordance with The British Psychological Society (BPS) code of human research ethics (see BPS, 2014, for review) as well as provided with a debrief sheet to take home (see Appendix H, for morning / afternoon order; and Appendix P, for afternoon / morning order).

5.3.5: Design

The experiment employed a 2 × 2 × 10 mixed between-within subjects analysis of variance (ANOVA) with order effect (i.e., Group 1: afternoon/morning or Group 2: morning/afternoon) as the between-subjects factor, × time of day (morning or afternoon) × time on task (1-minute, 2-minutes, 3-minutes, 4-minutes, 5-minutes, 6-minutes, 7-minutes, 8-minutes, 9-minutes or 10-minutes) as the within-subjects factors. The morning session (i.e., before 11:00) and afternoon session (i.e., after 17:00) were held on two consecutive days.

5.3.6: Procedure

In order to ensure participants were fully aware of the inclusion and exclusion criteria, all participants were contacted using Cardiff University's Experimental Management System (EMS) emailing system 48 hours prior participation and further reminded 24 hours prior participation in addition to being provided with brief instructions through EMS.

The study was administered using Apple's iPhone 6s Plus as previous research had identified that participant's reaction times (RTs) from Apple's mobile devices were more in line with those found in the literature (see Evans *et al.*, 2019, for review). In order to ensure no order effect, participants were randomly assigned to one of two counterbalanced groups, with each group requiring that participants come in on two consecutive days. Group 1 were instructed to come in for the afternoon session (i.e., after 17:00) first, followed by the morning session (i.e., before 11:00) on the second day. Group 2 were instructed vice versa. This study consisted of two parts. The first part was the online 10-minute mobile Psychomotor Vigilance Task (online 10-min m-PVT) reaction time test, which was a modified version of the Dinges and Powell's (1985) Psychomotor Vigilance Task. The m-PVT was run on the Dolphin Web Browser mobile application. The second part was the demographic questionnaire that was distributed within Qualtrics Surveys mobile application. On the online 10-min m-PVT (see Figure 62), participants were presented with on-screen instructions and a button at the end that read 'Start'. In each trial, participants were shown a black screen background, and at the centre of the screen they would be presented with a large red fixation circle. The red fixation circle (i.e., inter-stimulus interval) would remain on the screen for a randomised duration that lasted between 2 – 10 seconds, which was then followed by a yellow stimulus counter. As soon as the inter-stimulus interval reached the randomised duration, a yellow stimulus counter appeared counting up in milliseconds from 0 – 5 seconds where it would lapse (i.e., error of omission for 0.5 seconds) and begin the next trial, or until the participant tapped on the screen. Once the participant tapped on the screen, their reaction time (i.e., stimulus) would be displayed for 0.5 seconds. At the end of each trial, a black background would appear on-screen for 0.5 seconds. There were 92 trials in total that lasted approximately 10 minutes. Kribbs and Dinges (1994) found that after a maximum of three trials, the practice effect for the PVT was removed. This study conservatively implemented five practice trials to ensure participants were fully aware of the task, which were removed from final analyses. If participants responded prematurely during any trial (i.e., before the timer commenced counting up), the trial would reset. To also ensure participants were made aware of their premature response, the following message in red was displayed on the centre of the screen, 'You clicked too early! This trial will be reset.' A visual illustration of the online 10-minute mobile Psychomotor Vigilance Task (online 10-min m-PVT) is presented in Figure 63.

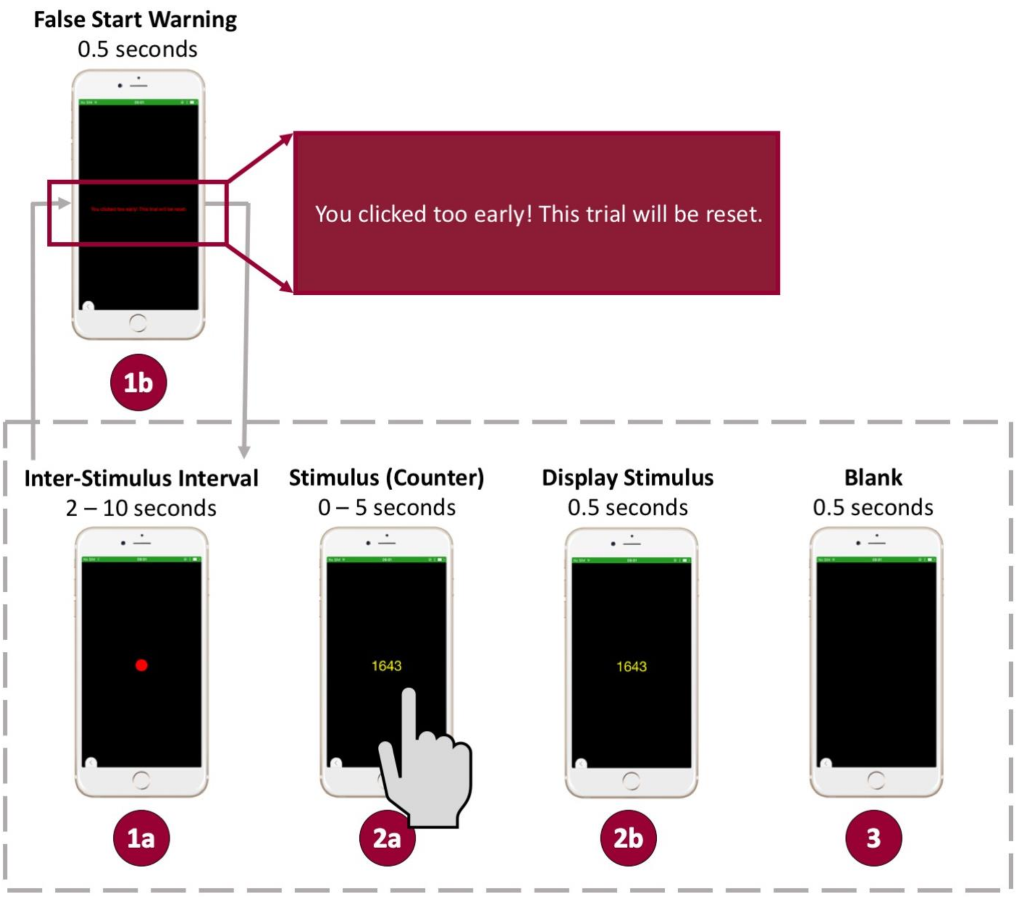


Figure 62: Mobile Psychomotor Vigilance Task (m-PVT) timeline

1a. Participants were presented with a large red circle (i.e., inter-stimulus interval), which appeared for a randomised duration between 2 – 10 seconds.

1b. If participants responded prematurely, a false start warning message appeared informing them that they clicked too early and that the trial would be reset.

2a. As soon as the inter-stimulus interval reached the randomised duration, a yellow stimulus counter appeared counting up in milliseconds from 0 – 5 seconds where it would lapse (i.e., error of omission for 0.5 seconds) and begin the next trial, or until the participant had tapped on the screen.

2b. Once the participants had tapped on the screen, their reaction time (i.e., stimulus) would be displayed for 0.5 seconds.

3. At the end of each trial, a black background would appear on-screen for 0.5 seconds.

Figure 63: Visual illustration of the Mobile Psychomotor Vigilance Task (m-PVT)

5.4: Results

The aim of this study was to use the time-of-day and time-on-task effect to replicate and validate whether the alternative online mobile version of the 'gold standard' 10-minute Psychomotor Vigilance Task (10-min PVT) i.e., online 10-minute mobile Psychomotor Vigilance Task (online 10-min m-PVT) could be used to provide an objective indicator of fatigue in frontline safety critical workers. IBM's Statistical Package for the Social Sciences (SPSS) version 23 for Mac was used to analyse the data. A total of 11,904 test trials (i.e., 87 test trials per session) were submitted for data analyses, with all 737 practice trials (i.e., 5 practice trials per session) excluded from final analyses from all 12,641 recorded trials. It is important to note that all mobile devices running the online 10-min m-PVT were administered through the Dolphin internet browser and were connected using Cardiff University's eduroam Wi-Fi roaming service. Therefore, in rare occasions when the Wi-Fi connectivity dropped, participant's trials were lost and thus not recorded. As a result, a total of 7.71 per cent (n = 975) trails of all potential 13,616 trials (i.e., 740 practice and 12,876 test) were lost and not recorded. Due to the large variability in the reporting of the PVT outcome metrics (see Basner & Dinges, 2011, for review), this study adopted the three most frequently reported PVT outcome analyses; mean reaction time (RT), mean speed response (1/RT) and mean number of lapses, respectively. In addition, based on Basner and Dinges (2011) recommendations, all 11,904 test trials with reaction time (RTs) < 100 ms (i.e., false start), which accounted for .03 per cent (n = 4) and RTs ≥ 500 ms (i.e., number of lapses), which accounted for 6.42 per cent (n = 811), were considered for exclusion from the final mean reaction time (RT) and mean speed response (1/RT) analyses. All 6.42 per cent (n = 811) of RTs ≥ 500 ms (i.e., number of lapses) were analysed separately.

5.4.1: Reliability and validity

The following subsections address the reliability and validly of the online 10-min m-PVT by highlighting the importance of; reporting the effect size to provide a confidence level in the reported *p*-values, reporting the intraclass correlation coefficient (ICC), which is a widely used reliability analysis, and reporting Mauchly's Test of Sphericity, which tests the hypothesis that the variances of the differences between conditions are equal.

5.4.1.1: Effect size

According to Levine and Hullett (2002), statistical tests are highly dependent on sample size e.g., if sample size are too small, strong and important effects can be nonsignificant (i.e., a Type II error is made). In contrast, when sample sizes are too large, even trivial effects can have impressive looking *p*-values. Therefore, reporting the effect size is crucial as it provides an estimate of the magnitude of the effect that is relatively independent of the sample size. As a result, the effect size provides an additional level of confidence in the *p*-value. Keppel (1991) has recommended partial eta-squared (η_p^2) in order to be able to improve the compatibility of ANOVA effect size between studies, since eta-squared (η^2) cannot easily be comparable between studies (Lakens, 2013). As a result, the following η_p^2 threshold criteria were used; .01 small, .06 medium, and .14 large (Cohen, 1973; Cohen, 1988).

5.4.1.2: Reliability

Even though Dinges and Powell (1985) have shown that the 10-min PVT is highly reliable, Ko and Li (2016) outline that before any assessment tools or measurement instruments can be used for research or clinical applications, their reliability must be established. According to Daly and Bourke (2000), reliability is defined as the extent to which measurements can be replicated. However, Weir (2005) states that reliability refers to the consistency of a test or measurement. In contrast, DeVon *et al*. (2007) argue that reliability is defined as a measure of true scores and includes an examination of stability and equivalence. However, Weir (2005) acknowledges that the term 'reliability' has become a confusing definition for many researchers due to the jargon used in the context of reliability within the field of psychology i.e., consistency, precision, repeatability, and agreement. For example, even the term reliability that is conceptualised as consistency, tends to consist of both absolute consistency and relative consistency (see Safrit, 1975, for review). Historically, reliability has been evaluated with Pearson correlation coefficient, Bland-Altman plot, and paired *t*-test (Bland & Altman, 1986; Brown, Lucero, & Foss, 1962; Bruton, Conway, & Holgate, 2000; Hopkins, 2000). Ko and Li (2016) highlight that intraclass correlation coefficient (ICC) is a widely used reliability index in test-retest, intrarater, and interrater reliability analyses. Moreover, Ko and Li (2016) further outline that the selection of an appropriate ICC form for reliability analysis involves identification of the type of reliability study to be conducted, followed by determining the 'Model,' 'Type,' and 'Definition' selection to be used, since there are 10 forms of ICCs. Therefore, since the output metrics of the 10-min PVT generates output data that is continuous in nature due to it being reaction times, the 2-way mixed-effects model is an appropriate test for testing intrarater reliability with multiple scores from the same rater i.e., researcher (Shrout & Fleiss, 1979) as well as it being used in test-retest reliability research due to the fact that repeated measurements cannot be regarded as randomised samples (Portney & Watkins, 2000). Furthermore, it is recommended that absolute agreement definition should always be chosen for both test-retest and intrarater reliability research due to the fact that measurements would be meaningless if there is no agreement between repeated measurements (Ko & Li, 2016). As a result, interpretations are as following; values less than 0.5 are indicative of 'poor' reliability, values between 0.5 and 0.75 indicate 'moderate' reliability, values between 0.75 and 0.9 indicate 'good' reliability, and values greater than 0.90 indicate 'excellent' reliability (Ko & Li, 2016).

Intraclass correlation coefficient (ICC) estimates and their 95% confident intervals were calculated using IBM's Statistical Package for the Social Sciences (SPSS) version 23 for Mac based on a mean-rating (k = 20), absolute-agreement, 2-way mixed-effects model. An excellent degree of reliability (see Ko & Li, 2016, for review) was found between all 20 mean reaction times (RTs) i.e., the morning online 10-min m-PVT (1-minute, 2 minutes, 3 minutes, 4 minutes, 5 minutes, 6 minutes, 7 minutes, 8 minutes, 9 minutes and 10 minutes) and the afternoon online 10-min m-PVT (1-minute, 2 minutes, 3 minutes, 4 minutes, 5 minutes, 6 minutes, 7 minutes, 8 minutes, 9 minutes and 10 minutes) measurements. Analysis revealed that the average measure of ICC was .960 with a 95% confidence interval from .944 to .972, $F(73, 1387) = 28.22$, $p < .001$.

5.4.2: Mauchly's Test of Sphericity for mean reaction time (RT)

One of the repeated measures ANOVA assumptions is that it requires sphericity. According to Field (2013), sphericity can be assessed using Mauchly's test, which tests the hypothesis that the variances of the differences between conditions are equal. As a result, if Mauchly's test statistic is significant, we conclude that there are significant differences between the variances of differences and, therefore, the condition of sphericity is not met. If, however, Mauchly's test statistic is non-significant, then it is reasonable to conclude that the variances of differences are roughly equal. In summary, if Mauchly's test is significant then we must be wary of the resulting F-ratios. Therefore, the degrees of freedom are adjusted using estimates of sphericity advocated by Greenhouse and Geisser (1959) and Huynh and Feldt (1976). Many authors recommend that when estimates of sphericity are greater than .75 the Huynh–Feldt estimate should be used, but when the Greenhouse–Geisser estimate of sphericity is less than .75 or nothing is known about sphericity at all the Greenhouse–Geisser correction should be used (Barcikowski & Robey, 1984; Girden, 1992; Huynh & Feldt, 1976). For the mean reaction time (RT) analyses, Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated for the time on task, $\chi^2(44) = 73.95$, $p = .003$. Therefore, the Huynh-Feldt test was reported instead of Sphericity Assumed since Greenhouse–Geisser Epsilon was greater than .75 for the time on task, and Greenhouse–Geisser test was reported for time of day as nothing was known about sphericity. All other Mauchly's Test of Sphericity indicated that the assumption of sphericity had not been violated i.e., time of day × time on task, $\chi^2(44) = 58.44$, $p = .072$.

5.4.3: Mean reaction time (RT)

Figure 64 presents the illustrated mean reaction times (RTs) across the different conditions. All RTs (i.e., RTs >100 ms and < 500 ms) were submitted to a 2 × 2 × 10 mixed between-within subjects analysis of variance (ANOVA) with order effect (i.e., Group 1: afternoon/morning or Group 2: morning/afternoon) as the between-subjects factor, × time of day (morning or afternoon) × time on task (1-minute, 2 minutes, 3 minutes, 4 minutes, 5 minutes, 6 minutes, 7 minutes, 8 minutes, 9 minutes or 10 minutes) as the within-subjects factors.

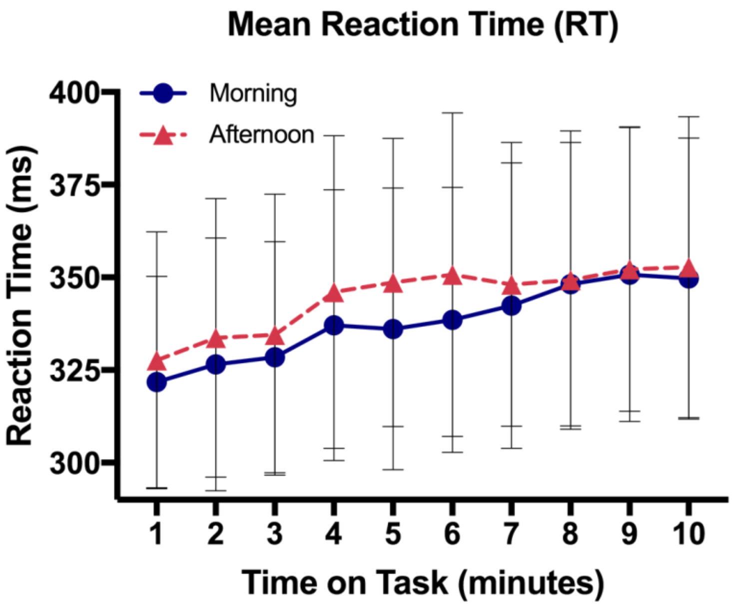


Figure 64: Mean reaction times (RTs) for time of day and time on task of the mobile Psychomotor Vigilance Task (m-PVT)

Note: Mean speed responses (1/RTs) and Mean reaction times (RTs), respectively, are presented in bins of 1-minute intervals. Error bars represents standard deviations.

There was no significant main effect comparing the two counterbalanced groups (i.e., Group 1: afternoon/morning or Group 2: morning/afternoon), $F(1, 72)$, 3.56, $p = .063$, $\eta_p^2 = .047$. There was a significant main effect of time of day, Greenhouse–Geisser = 1.00, $F(1, 72)$, 5.59, $p = .021$, $\eta_p^2 = .07$, indicating a medium effect size (Cohen, 1973; Cohen, 1988; Levine & Hullett, 2002), which provides a moderate level of confidence in the reported p-value. Furthermore, there was also a significant main effect of time on task, Huynh-Feldt = .908, $F(8.17, 588.18)$, 25.78, $p < .001$, $\eta_p^2 = .26$, indicating a large effect size (Cohen, 1973; Cohen, 1988; Levine & Hullett, 2002), which provides a high level of confidence in the reported p-value. All interactions were not significant. There was no significant: two-way interaction, time of day × time on task, Sphericity Assumed, F(9, 648), 1.24, p = .265, η_p^2 = .02; two-way interaction, order effect × time of day, Sphericity Assumed, F(1, 72), .455, p = .502, η_p^2 = .006; two-way interaction, order effect × time on task, Sphericity Assumed, F(9, 648), .701, p = .708, η_p^2 = .01;and three-way interaction, order effect × time of day × time on task, Sphericity Assumed, F(9, 648), .657, p = .748, η_p^2 = .009.

5.4.3.1: Mean reaction time (RT) Post-Hoc tests

The main effect of time of day for mean reaction times (RTs) is illustrated in Figure 65 and was followed by post-hoc tests with Bonferroni correction for multiple comparisons. Post-hoc tests also showed that participants mean RTs were also significantly faster in the morning session (M = 337.94 ms, SE = 3.38 ms) when compared to RTs in the afternoon session (M = 344.36 ms, SE = 3.80 ms) p = .021.

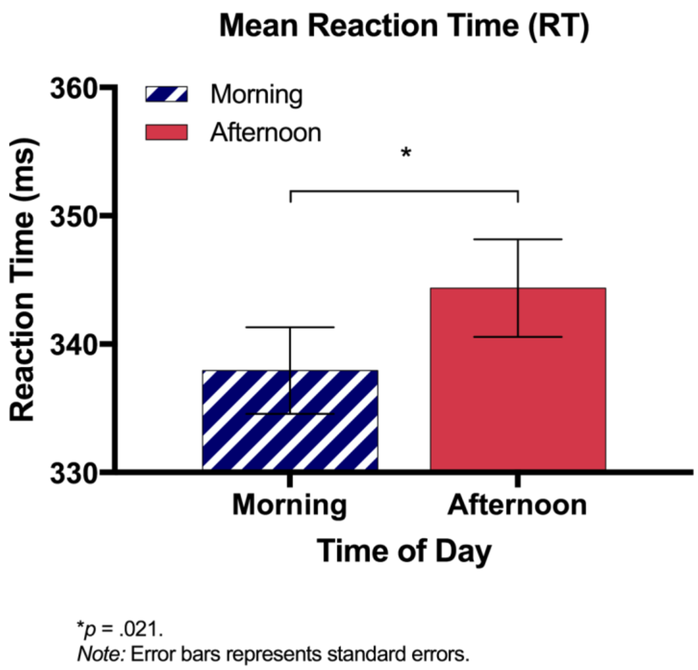
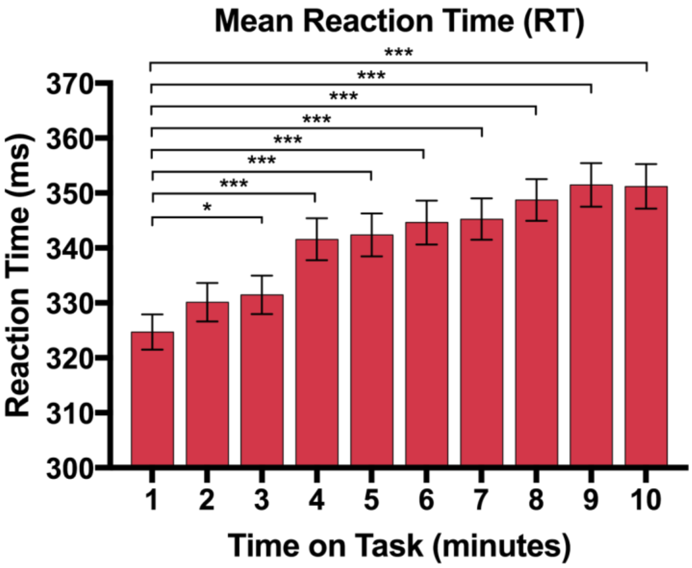


Figure 65: Mean reaction times (RTs) for time of day (morning or afternoon)

The main effect of time on task for the mean reaction times (RTs) is illustrated in Figure 66 and was followed by post-hoc tests with Bonferroni correction for multiple comparisons. Post-hoc t-tests showed that participants mean RTs were significantly slower after 3 minutes (*M* = 331.48 ms, *SE* = 3.50 ms) when compared to RTs after 1-minute (*M* = 324.69 ms, *SE* = 3.21 ms) $p = .036$, on the m-PVT temporal attention task. In addition, participants mean RTs were significantly slower thereafter from the first minute when compared to; 4 minutes ($p < .001$), 5 minutes ($p < .001$), 6 minutes ($p < .001$), 7 minutes ($p < .001$), 8 minutes ($p < .001$), 9 minutes ($p < .001$) and 10 minutes ($p < .001$). There was no significant difference when comparing mean RTs from the first minute (*M* = 324.69 ms, *SE* = 3.21 ms) and second minute (*M* = 330.11 ms, *SE* = 3.50 ms) $p = .854$.



*p < .05, **p < .01, ***p < .001.
Note: Mean RTs are presented in bins of 1-minute intervals. Error bars represents standard errors.

Figure 66: Mean reaction times (RTs) for time on task

5.4.4: Mauchly's Test of Sphericity for mean speed response (1/RT)

The assumption of sphericity was examined using the Mauchly's Test of Sphericity. For the mean speed response (1/RT) analyses, Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated for the time on task, $\chi^2(44) = 63.61$, $p = .028$, and two-way interaction, time of day × time on task, $\chi^2(44) = 61.79$, $p = .040$. Therefore, the Huynh-Feldt test was reported instead of Sphericity Assumed since Greenhouse–Geisser Epsilon was greater than .75 for the time on task, and time of day × time on task.

In addition, Greenhouse–Geisser test was reported for time of day as nothing was known about sphericity.

5.4.5: Mean speed response (1/RT)

Figure 67 presents the illustrated mean speed responses (1/RTs) across the different conditions. All 1/RTs (i.e., 1/RTs >100 ms and < 500 ms) were submitted to a 2 × 2 × 10 mixed between-within subjects analysis of variance (ANOVA) with order effect (i.e., Group 1: afternoon/morning or Group 2: morning/afternoon) as the between-subjects factor, × time of day (morning or afternoon) × time on task (1-minute, 2 minutes, 3 minutes, 4 minutes, 5 minutes, 6 minutes, 7 minutes, 8 minutes, 9 minutes or 10 minutes) as the within-subjects factors.

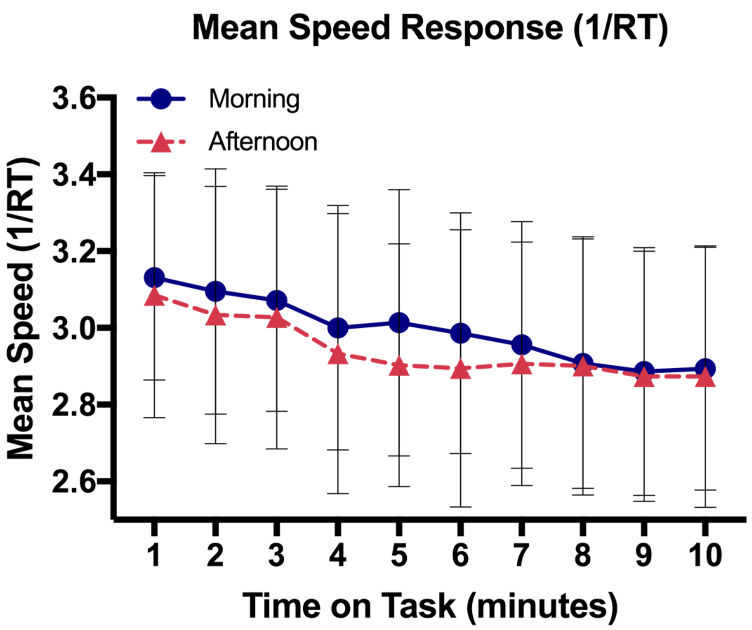


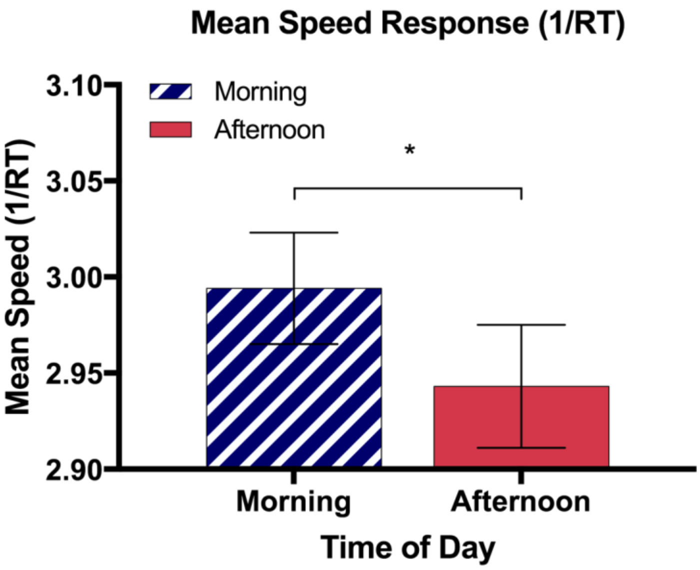
Figure 67: Mean speed responses (1/RTs) for time of day and time on task of the mobile Psychomotor Vigilance Task (m-PVT)

Note: Mean speed responses (1/RTs) and Mean reaction times (RTs), respectively, are presented in bins of 1-minute intervals. Error bars represents standard deviations.

There was no significant main effect comparing the two counterbalanced groups (i.e., Group 1: afternoon/morning or Group 2: morning/afternoon), $F(1, 72)$, 3.52, p = .065, η_p^2 = .047. There was a significant main effect of time of day, Greenhouse–Geisser = 1.00, $F(1, 72)$, 4.54, p = .037, η_p^2 = .06, indicating a medium effect size (Cohen, 1973; Cohen, 1988; Levine & Hullett, 2002), which provides a moderate level of confidence in the reported *p*-value. Furthermore, there was also a significant main effect of time on task, Huynh-Feldt = 994, $F(8.49, 611.50)$, 24.42, $p < .001$, $\eta_p^2 = .25$, indicating a large effect size (Cohen, 1973; Cohen, 1988; Levine & Hullett, 2002), which provides a high level of confidence in the reported *p*-value. All interactions were not significant. There was no significant: two-way interaction, time of day × time on task, Huynh-Feldt = 959, $F(8.63, 621.44)$, 1.13, $p = .337$, $\eta_p^2 = .02$; two-way interaction, order effect × time of day, Sphericity Assumed, $F(1, 72)$, .296, $p = .588$, $\eta_p^2 = .004$; two-way interaction, order effect × time on task, Sphericity Assumed, $F(9, 648)$, .923, $p = .505$, $\eta_p^2 = .01$; and three-way interaction, order effect × time of day × time on task, Sphericity Assumed, $F(9, 648)$, .454, $p = .905$, $\eta_p^2 = .01$.

5.4.5.1: Mean speed response (1/RT) Post-Hoc tests

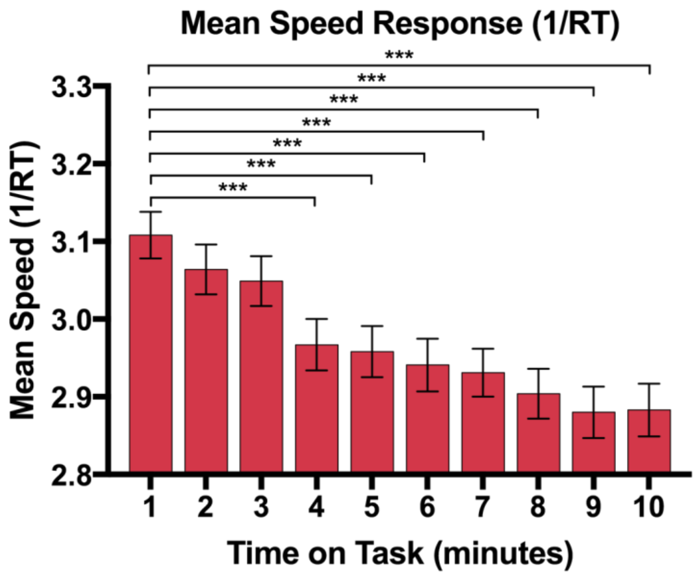
The main effect of time of day for mean speed responses (1/RTs) is illustrated in Figure 68 and was followed by post-hoc tests with Bonferroni correction for multiple comparisons. Post-hoc tests showed that participants mean 1/RTs were significantly higher in the morning session ($M = 2.99$, $SE = .03$) when compared to mean 1/RTs in the afternoon session ($M = 2.94$, $SE = .03$) $p = .037$.



*p = .037.
Note: Error bars represents standard errors.

Figure 68: Mean speed responses (1/RTs) for time of day (morning or afternoon)

The main effect of time on task for the mean speed responses (1/RTs) is illustrated in Figure 69 and was followed by post-hoc t-tests with Bonferroni correction for multiple comparisons. Post-hoc tests showed that participants mean 1/RTs were significantly slower in their speed responses after 4 minutes (M = 2.97, SE = .03) when compared to mean 1/RTs after 1-minute (M = 3.11, SE = .03) p < .001, on the m-PVT temporal attention task. In addition, participants mean 1/RTs were significantly slower in their speed responses thereafter from the first minute when compared to; 5 minutes (p < .001), 6 minutes (p < .001), 7 minutes (p < .001), 8 minutes (p < .001), 9 minutes (p < .001) and 10 minutes (p < .001). There was no significant difference in their speed responses when comparing the first minute (M = 3.11, SE = .03) and second minute (M = 3.06, SE = .03) p = 1.00, as well as when comparing the first minute (M = 3.11, SE = .03) and third minute (M = 3.05, SE = .03) p = .122.



*p < .05, **p < .01, ***p < .001.
Note: Mean 1/RTs are presented in bins of 1-minute intervals. Error bars represents standard errors.

Figure 69: Mean speed responses (1/RTs) for time on task

5.4.6: Mauchly's Test of Sphericity for mean number of lapses

The assumption of sphericity was examined using the Mauchly's Test of Sphericity. For the mean number of lapses analyses, Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated for the time on task, $\chi^2(44) = 124.19$, $p < .001$, and two-way interaction, time of day × time on task, $\chi^2(44) = 84.88$, $p < .001$. Therefore, the Huynh-Feldt test were reported instead of Sphericity Assumed since Greenhouse–Geisser Epsilon was greater than .75 for the time on task, and time of day × time on task. In addition, Greenhouse–Geisser test was reported for time of day as nothing was known about sphericity.

5.4.7: Mean number of lapses

From all test trials, a total of 6.42 per cent (n = 811) RTs ≥ 500 ms were submitted for data analyses. Figure 70 presents the illustrated mean number of lapses across the different conditions. The mean number of lapses were submitted to a 2 × 2 × 10 mixed between-within subjects analysis of variance (ANOVA) with order effect (i.e., Group 1:

afternoon/morning or Group 2: morning/afternoon) as the between-subjects factor, × time of day (morning or afternoon) × time on task (1-minute, 2 minutes, 3 minutes, 4 minutes, 5 minutes, 6 minutes, 7 minutes, 8 minutes, 9 minutes or 10 minutes) as the within-subjects factors.

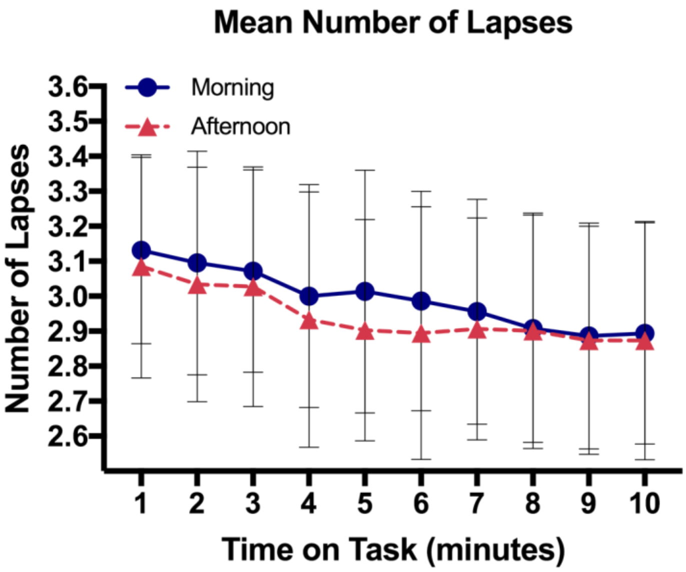


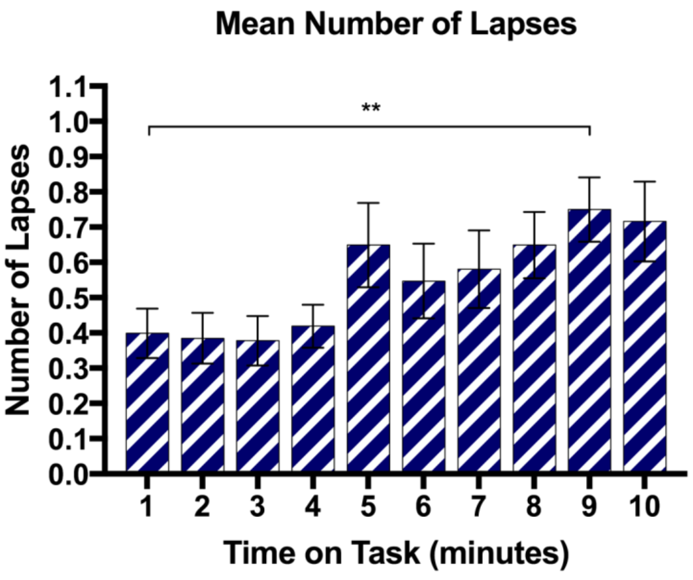
Figure 70: Mean number of lapses for time of day and time on task of the mobile Psychomotor Vigilance Task (m-PVT)

Note: Mean number of lapses are presented in bins of 1-minute intervals. Error bars represents standard deviations.

There was no significant difference when comparing the main effect of the two counterbalanced groups (i.e., Group 1: afternoon/morning or Group 2: morning/afternoon), $F(1, 72)$, .728, $p = .396$, $\eta_p^2 = .010$. There was a significant main effect of time on task, Huynh-Feldt = .772, $F(6.95, 500.23)$, 5.19, $p < .001$, $\eta_p^2 = .07$, indicating a medium effect size (Cohen, 1973; Cohen, 1988; Levine & Hullett, 2002), which provides a moderate level of confidence in the reported p-value. All other main effect and interactions were not significant: time of day, Greenhouse–Geisser = 1.00, $F(1, 72)$, .155, $p = .695$, $\eta_p^2 = .002$; two-way interaction, time of day × time on task, Huynh-Feldt = 882, $F(7.94, 571.65)$, 1.27, $p = .255$, $\eta_p^2 = .02$; two-way interaction, order effect × time of day, Sphericity Assumed, $F(1, 72)$, .242, $p = .625$, $\eta_p^2 = .003$; two-way interaction, order effect × time on task, Sphericity Assumed, $F(9, 648)$, 1.00, $p = .437$,

$\eta_p^2 = .01$; and three-way interaction, order effect × time of day × time on task, Sphericity Assumed, $F(9, 648)$, 1.57, $p = .121$, $\eta_p^2 = .02$.

The main effect of time on task for the mean number of lapses is illustrated in Figure 71 and was followed by post-hoc tests with Bonferroni correction for multiple comparisons. Post-hoc tests showed that participants mean number of lapses were only significantly different after 9 minutes ($M = .75$, $SE = .09$) when compared to the mean number of lapses after 1-minute ($M = .40$, $SE = .07$) $p = .003$, on the m-PVT temporal attention task. All other post-hoc tests with Bonferroni correction for multiple comparisons were not significant when comparing the first minute ($M = .40$, $SE = .07$) to; 2 minutes ($p = 1.00$), 3 minutes ($p = 1.00$), 4 minutes ($p = 1.00$), 5 minutes ($p = 1.00$), 6 minutes ($p = 1.00$), 7 minutes ($p = 1.00$), 8 minutes ($p = .480$) and 10 minutes ($p = .118$).



*p < .05, **p < .01, ***p < .001.
Note: Mean number of lapses are presented in bins of 1-minute intervals. Error bars represents standard errors.

Figure 71: Mean number of lapses

5.5: Discussion

The aim of this study using the time-of-day and time-on-task effect was to replicate and validate whether the alternative online mobile version of the 'gold standard' 10-minute

Psychomotor Vigilance Task (10-min PVT) i.e., online 10-minute mobile Psychomotor Vigilance Task (online 10-min m-PVT) could be used to provide an objective indicator of fatigue in frontline safety critical workers. Firstly, there was no significant main effect comparing the two counterbalanced groups (i.e., Group 1: afternoon/morning or Group 2: morning/afternoon), which seems to indicate there were no significant reaction time differences between the two counterbalanced groups. This is important to note since any differences could significantly impact interpretations of the findings. Secondly, intraclass correlation coefficient (ICC) analysis revealed an excellent degree of reliability in this study (see Ko & Li, 2016, for review) across all 20 mean reaction Times (RTs) which indicates that the online 10-min m-PVT was able to generate highly reliable RTs across all participants.

In this study it was found that the online 10-min m-PVT was sensitive enough to detect levels of fatigue using the time-of-day effect (i.e., morning versus afternoon), whereby participants had significantly slower reaction times (RTs) and greater speed responses (1/RTs) in the afternoon session when compared to the morning session. These findings also have a higher level of confidence as all significant *p*-values comprised of either a medium or high partial eta-squared (η_p^2) effect size. These results support Lenne *et al.*'s (1997) findings that performance is subject to diurnal variations as well as work by Carrier and Monk (2000) which outlined that human performance efficiency changes as a function of time of day. Furthermore, Smith (1992) had previously identified that performance was less accurate in the early evening when compared to the early morning. However, Smith (1992) also highlights that performance was also faster in the early evening when compared to the early morning. This is further supported by Lafrance and Dumont (2000) who examined diurnal variation in mental and physical performance in university football students and found that self-reported alertness ratings increased, while self-reported fatigue ratings decreased. However, Lambourne and Tomporowski (2010) argues that following exercise, cognitive task performance improved. Nevertheless, Lambourne and Tomporowski (2010) point out that there is a complex relation between exercise and cognition whereby cognitive performance may be impaired, enhanced or dependant on various factors such as; the time it is measured e.g., immediately post exercise or prolonged post exercise, the type of cognitive task e.g., simple reaction time task (i.e., PVT), as well as the type of exercise that is being performed e.g., football, rugby, running, cycling, karate, etc. In contrast, Patkai (1971) found that performance does not vary during the day for morning workers, while evening workers showed a steady improvement. However, it is also important to acknowledge that these findings could also be explained by participants eating and/or drinking before the morning session, since it is highly likely that participants would have had ample opportunities for breakfast before commencing the study at 11:00 when compared to the opportunity to get food prior to the afternoon session i.e., 17:00.

Analyses from the number of lapses i.e., RTs ≥ 500 ms did not reveal any additional information, which is not surprising when we consider that there was only a total of 6.4 per cent number of lapses recorded. However, Basner and Dinges (2011) reviewed journal manuscripts that reported PVT outcome metrics and identified that the most commonly reported PVT outcome was number of lapses, which represented over 60 per cent of all 141 journals examined. In contrast, this study recruited undergraduate psychology students. Therefore, perhaps in this cohort, students' cognitive neurobehavioral performances (i.e., reactions) in an educational environment setting such as a university may not provide sufficient sustained mental workload to elicit the time-of-day effect (Carrier & Monk, 2000; Lafrance & Dumont, 2000; Lenne *et al.*, 1997; Patkai, 1971; Smith, 1992) to detect fatigue levels. Nevertheless, findings from this study are in line with PVT outcome metrics found in Evans *et al.* (2019) in terms of what was expected from using an Apple iPhone 6s Plus mobile device. In summary, Evans *et al.* (2019)

identified that the alternative online mobile version of the PVT generated reaction times that were either more suited for reaction time outcome analyses (i.e., RT range: 100 ms – 499 ms) or number of lapses outcome analyses (i.e., RTs ≥ 500 ms). These differences could have been introduced as a result of both manufacture hardware configurations (i.e., processing power, display size, display refresh rate, pixel density, etc.) as well as software differences (i.e., operating system, the programming language used to build the app, efficiency of the developers' programming code, etc.).

Findings from this study also support previous research that have identified an increase in fatigue results in impaired alertness (Dorrian *et al.*, 2000; Gillberg *et al.*, 1994), whereby time-on-task effect i.e., sustained attention (see Langner & Eickhoff, 2013; Mackworth, 1948; Mackworth, 1968, for reviews), as measured in RTs and 1/RTs significantly reduces after 4 minutes of continuous performance using the online 10-min m-PVT. These findings are consistent with previous work, which suggested that sustained attention drops with prolonged duration of task (Dinges & Powell, 1988; Dinges & Powell, 1989; Doran, van Dongen, & Dinges, 2001; Langner & Eickhoff, 2013; Mackworth, 1948; Mackworth, 1968). Therefore, this study seems to also suggest that an online mobile version of the 'gold standard' 10-min PVT could be used to objectively measure sensitivity levels of fatigue after 4 minutes on the online 10-min m-PVT for frontline safety critical settings e.g., train drivers, hospital staff, emergency services, law enforcers, etc. However, this study was not without limitations.

Despite studies outlining the availability to run online experiments (Barnhoorn *et al.*, 2015; Crump *et al.*, 2013; Reimers & Stewart, 2016) and other studies demonstrating comparability between lab based and online based experiments (e.g., de Leeuw & Motz, 2016; Reimers & Stewart, 2007; Schubert *et al.*, 2013; Simcox & Fiez, 2014), in rare occasions when the Wi-Fi connectivity dropped, participant's trials were lost and thus not recorded. As a result, a total of 7.71 per cent (n = 975) of test trials were never recorded.

Therefore, there was a clear need to develop an offline iOS mobile app version of the online 10-min m-PVT i.e. offline 10-min m-PVT. However, an iOS mobile app would require careful consideration and significant investment of time and resources for the various phases of the iOS mobile app process; pre-production, design, development, and launch. Nevertheless, before such an ambitious project is undertaken, further research is needed to determine whether a shorter mobile version of the online 10-minute mobile Psychomotor Vigilance Task (online 10-min m-PVT) i.e., online 5-min m-PVT could be used to provide an objective indicator of simulated workload fatigue.

5.6: Chapter summary

The aim of this chapter was to use the time-of-day and time-on-task effect to replicate and validate whether the alternative online mobile version of the 'gold standard' 10-minute Psychomotor Vigilance Task (10-min PVT) i.e., online 10-minute mobile Psychomotor Vigilance Task (online 10-min m-PVT) could be used to provide an objective indicator of fatigue for frontline safety critical workers such as train drivers, hospital staff, emergency services, law enforcers, etc. 74 (65 female, M = 19.41, SD = 1.61) participants were recruited voluntarily from Cardiff University via the Experimental Management System (EMS) to take part in the 10-min m-PVT. The study involved participants attending two sessions, a morning session (i.e., before 11:00) and an afternoon session (i.e., after 17:00), which were held on two counterbalanced consecutive days. This study found a significant main effect of time of day, whereby both mean reaction times (RTs) and mean speed response (i.e., reciprocal of reaction time) (1/RT) significantly differed between the morning session and afternoon session. Post-hoc tests showed that participants mean RTs were significantly faster in the morning session when compared to mean RTs in the afternoon session. Post-hoc tests also showed that participants mean 1/RTs were significantly slower in the morning session when compared to mean 1/RTs in the afternoon session. This study also found a significant main effect of time on task for both mean RTs and mean 1/RTs. Post-hoc tests also showed that participants mean RTs were significantly slower after 3 minutes when compared to mean RTs after 1-minute on the m-PVT temporal attention task. In addition, post-hoc tests also showed that participants mean 1/RTs were significantly faster after 4 minutes when compared to their mean 1/RTs after 1-minute on the online m-PVT temporal attention task. These findings do seem to indicate that an alternative and shorter online mobile version of the 'gold standard' 10-minute Psychomotor Vigilance Task (10-min PVT) i.e., online 10-minute mobile Psychomotor Vigilance Task (online 10-min m-PVT) could be used to provide an objective indicator of fatigue for frontline safety critical workers such as train drivers, hospital staff, emergency services, law enforcers, etc. Further research is now needed to determine whether a shorter mobile version of the online 10-min m-PVT i.e., online 5-min m-PVT could be used to provide an objective indicator of simulated workload fatigue.

Chapter 6: Investigating a shorter mobile version of the online 10-min m-PVT i.e., online 5-min m-PVT as an objective indicator of simulated workload fatigue

6.1: Overview of chapter

In the previous chapter it was identified that the alternative online mobile Psychomotor Vigilance Task (online 10-min m-PVT) was sensitive enough to detect levels of fatigue between the morning and afternoon session i.e., time-of-day effect as well as the online 10-min m-PVT being able to detect levels of fatigue after only 4 minutes on the task i.e., time-on-task effect. Therefore, the aim of this study is now to investigate whether a shorter mobile version of the online 10-min m-PVT i.e., online 5-min m-PVT could be used to provide an objective indicator of simulated workload fatigue.

6.2: Introduction and rationale

In order to be able to meet task demands, there is usually a required amount of operator resources needed, referred to as human mental workload (Eggemeier, Wilson, Kramer, & Damos, 1991). According to Hart and Staveland (1988), human mental workload can be defined as a 'cost incurred by a human operator to achieve a particular level of performance and evolves from interactions between task demands, circumstances, skills, behaviour, and perceptions.' Therefore, human mental workload – often referred to as cognitive load – can be intuitively defined as the amount of mental work necessary for a person to complete a task over a given period of time (Longo, 2015; Longo, 2017). However, nowadays human mental workload is more generally defined as the measurement of the amount of mental resources involved in a cognitive task (Zammouri, Moussa, & Mebrouk, 2018). Grech *et al*. (2009) argue that the relationship between workload and fatigue changes over consecutive days. In their study they found that at the beginning of a 14-day period, low workload was associated with fatigue. In contrast, at the end of the 14-day period, high workload was associated with fatigue. Therefore, Gui *et al*. (2015) outline that after prolonged mental workload, individuals will begin to show reduced behavioural performance as well as increased self-reported levels of fatigue, which are more commonly known as 'time-on-task' effects.

Sustaining attention to a demanding cognitive task will often come at a cost, which is known in the literature as the 'vigilance decrement,' or 'time-on-task effect' (Langner & Eickhoff, 2013; Mackworth, 1948; Mackworth, 1968). Gunzelmann, Moore, Gluck, van Dongen and Dinges (2011) highlights that observed vigilance tasks performance progressively degrades as the duration of those tasks increases. Satterfield, Wisor, Schmidt and van Dongen (2017) state that the Psychomotor Vigilance Task (PVT) is a vigilance task that has a well-documented time-on-task effect as well as the ability to generate a steady increase in the standard deviation of response times (RTs) across the standard 10-min PVT duration of the task (Doran, van Dongen, & Dinges, 2001). Warm *et al*. (2008) further argue that the time-on-task effect arises due to the fact that workload associated with tasks that require continuous vigilance consumes significantly more mental resources. Therefore, the use of vigilance tasks cannot immediately replenish mental resources.

Fatigue due to mental workload can be measured using a variety of psychological and physiological techniques, which include; subjective psychological self-reported measures e.g., the NASA Task Load Index (NASA-TLX) (Byrne *et al.*, 2010; Hart & Staveland, 1988; Orlandi & Brooks, 2018; Shakouri, Ikuma, Aghazadeh, & Nahmens, 2018) and the NASA-MATB (National Aeronautics and Space Administration Multi-Attribute Task Battery (Comstock & Arnegard, 1992) as well as objective physiological measures e.g., heart rate (HR) (e.g., Shakouri, Ikuma, Aghazadeh, & Nahmens, 2018), galvanic skin response (GSR) (e.g., Widyanti, Hanna., Muslim, & Sutalaksana, 2017), body temperature (e.g., Vergara, Moenne-Loccoz, & Maldonado, 2017), electrocardiogram (ECG) (e.g., Heine, Lenis, Reichensperger, Beran, Doessel, & Deml, 2017), electroencephalogram (EEG) (e.g., Berka *et al*., 2007; Hogervorst, Brouwer, & Van Erp, 2014; Hsu, Wang, Chen, & Chen, 2015; Jimenez-Molina, Retamal, & Lira, 2018; Shaw *et al*., 2018; So, Wong, Mak, & Chan, 2017), and eye tracking (Saito, 1992; Xu, Min, & Hu, 2018), and which have been extensively examined in various safety critical environments including; aviation (Blanco *et al*., 2018; Borghini, Astolfi, Vecchiato, Mattia, & Babiloni, 2014; Orlandi & Brooks, 2018), train driving (Myrtek *et al*., 1994), vehicle driving (Brookhuis & De Waard, 2001; Foy & Chapman, 2018; Paxion, Galy, & Berthelon, 2014), and in an operating theatre (Byrne *et al.*, 2010) but to name a few.

Landrigan *et al*. (2004) identified using the PVT that intensive care unit interns made substantially more serious medical errors when they worked frequent shifts of 24 hours when compared to working shorter shifts. In addition, Arnedt, Owens, Crouch, Stahl and Carskadon (2005) evaluated the association between subjective measures and objective measures of simulated driving performance and identified significantly impaired performance and vigilance i.e., sustained attention. Moreover, Gui *et al*. (2015) investigated the time-on-task effect by administering a 20-min PVT and found that participants exhibited slower reaction times (RTs) and more lapses at the end of the PVT than at the beginning. In addition, it was also found that participants reported greater mental fatigue ratings after completing the task (Gui *et al*., 2015). However, Lim *et al*. (2010) induced time-on-task effect by asking subjects to perform a 20-min PVT and found that participants' reaction times (RTs) were significantly slower as the task proceeded as well as participants subjectively self-reported higher fatigue ratings after the task. Moreover, it was also found that there was substantial inter-individual variation in participant's RTs by the end of the task. As a result, no significant associations were observed between self-reported mental fatigue changes and objective performance decline. This is further supported by Lim *et al*. (2010) and Parasuraman *et al*. (2009) who have outlined that not all individuals show the same extent of vulnerability to fatigue, as well as there being large inter-individual differences in responses to sleep loss and mental workload. As a result, in frontline safety critical settings, mental workload due to prolonged sustained attention i.e., time-on-task effect, are major contributing factors in traffic and occupational accidents (Bergasa, Nuevo, Sotelo, Barea, Lopez, 2006). Therefore, the time-on-task effect is a critical determinant of productivity and safety, most notably utilised in the transportation industry (Caldwell, 2005; Satterfield & van Dongen, 2013; Verster & Roth, 2013). Evirgen, Oniz and Ozgoren (2015) investigated the time-on-task effect by comparing two independent groups. Group one started by completing the PVT followed by a computerised 60-minute battery of tests used to assess various cognitive functions and then retaking the PVT (i.e., high workload condition). Group two also started by completing the PVT but instead of it being followed by a battery of various cognitive functional tests, a 40-minute resting period followed and then retaking the PVT (i.e., low workload condition). In their study it was found that participants who were in the high workload condition had significantly slower PVT RTs after completing 60-minute of various cognitive functional tests when compared to their first PVT RTs. In contrast, there were no significant difference found between participants first and second PVT, post the 40-minute resting period for the low workload condition. In addition, Goel, Abe, Braun and Dinges (2014) identified that high workload increased subjective self-reported fatigue and sleepiness ratings, regardless of sleep duration. However, it was also found that sleep restriction produced cumulative increases in the number of lapses, as measured by the PVT. Therefore, the aim of this study is to investigate whether a shorter mobile version of the online 10-minute mobile Psychomotor Vigilance Task (online 10-min m-PVT) i.e., online 5-min m-PVT could be used to provide an objective indicator of simulated workload fatigue.

6.3: Methodology

6.3.1: Ethical approval

The study received ethics approval from Cardiff University's Ethics Committee (EC.16.02.09.4464R4A4). The study conformed to the seventh amendment of the Declaration of Helsinki 1964 (World Medical Association, 2013) and was in accordance with the Data Protection Act 1998 and the General Data Protection Regulation (GDPR). All participants gave their Informed written consent (counterbalanced: high simulated workload / low simulated workload condition, Appendix Q; or low simulated workload / high simulated workload condition, Appendix CC) as well as electronic consent following the explanation of the nature of the study.

6.3.2: Participants

39 (11 male and 28 female) mean age 19.95 (SD = 2.15) participants were voluntarily recruited from Cardiff University via the Experimental Management System (EMS) to take part in the study. Participants were instructed not to consume caffeine (e.g., coffee, energy drinks, tea, etc.) and alcohol during the 24 hours before the study. The study involved participants attending a two-part study, that was administered using a mobile device. The two parts were completed on two different counterbalanced days, a HIGH simulated workload day and a LOW simulated workload day. A total of 10 credits were awarded through EMS, or £20 upon successful completion of all two-parts. Participation of all two-parts combined took approximately 120 minutes to complete.

6.3.3: Statistical analyses

IBM's Statistical Package for the Social Sciences (SPSS) version 23 for Mac was used to analyse the data. A combination of various statistical procedures were carried out on the data; intraclass correlation coefficient (ICC) analysis, descriptive analyses, mixed-design analysis of variance (ANOVA) and a two-way analysis of variance (ANOVA) to further explore interactions. The level of $\alpha < .05$ was used for all statistical tests of this experiment.

6.3.4: Materials and apparatus

The mobile Psychomotor Vigilance Task (m-PVT) was presented to participants on an Apple iPhone 6s Plus running Apple's iOS version 9.3.1 (Apple Inc.). The m-PVT ran in the following hardware configurations; system chip (Apple A9 APL1022), processor (Dual-core, 1840 MHz, Twister, 64-bit), graphics processor (PowerVR GT7600), and system memory (2048 MB RAM). The m-PVT was displayed on a 5.5-inch (diagonal) 1920 × 1080-pixel native resolution at 401 ppi Retina high definition display. The m-PVT was programmed using the client code HTML (hypertext markup language), and CSS (cascading style sheets) for the page visualisation and layout. JavaScript was also used to initiate the m-PVT, which was run using the Dolphin Web Browser (MoboTap Inc.) on the Apple's iPhone 6s Plus (Dolphin Web Browser version 9.9.0, released August 2011). The rational for selecting the Dolphin Web Browser for this study was that it allowed the full screen feature to be enabled, while other more native internet browsers did not, such as Safari, Chrome, and Firefox but to name a few. The online Qualtrics Surveys (Qualtrics Labs, Inc. version 13.28.06) was also used to collect demographic information. In order to increase validity and standardisation, all instructions were provided to participants in written form for both the counterbalanced conditions i.e., high simulated workload / low simulated workload (see Appendix R, S, T, U, V, W, X, Y, Z, and AA, for participant's user guides), or low simulated workload / high simulated workload (see Appendix DD, EE, FF, GG, HH, II, JJ, KK, LL, and MM, for participant's user guides), Participants were also verbally debriefed at the end of the study to explain the nature of the study in accordance with The British Psychological Society (BPS) code of human research ethics (see BPS, 2014, for review) as well as provided with a debrief sheet to take home i.e., high simulated workload / low simulated workload condition debrief (Appendix BB) or low simulated workload / high simulated workload condition debrief (Appendix NN).

6.3.5: Design

The experiment employed a 2 × 2 × 5 three-way repeated analysis of variance (ANOVA) comparing 2 × simulated workload (high simulated workload i.e., completing a 30-minutes battery of cognitive performance tasks or low simulated workload i.e., watching a television show) × 2 time of day (Time 1: pre-simulated workload m-PVT, or Time 2: post-simulated workload m-PVT) × 5 time on task (1-minute; 2 minutes; 3 minutes, 4 minutes; or 5 minutes).

6.3.6: Procedure

In order to ensure participants were fully aware of the inclusion and exclusion criteria, all participants were contacted using Cardiff University's Experimental Management System (EMS) emailing system 48 hours prior participation and further reminded 24 hours prior participation in addition to being provided with brief instructions through EMS. The study involved participants attending two sessions, each lasting approximately 60-minutes. The 5-min m-PVT was carried out both before the simulated workload (i.e., Time 1) and after the simulated workload (i.e., Time 2). The study involved participants attending two counterbalanced morning (i.e., 07:30am) sessions; a high simulated workload session and a low simulated workload session.

The study was administered using Apple's iPhone 6s Plus as previous research had identified that participant's mean reaction times (RTs) and mean speed responses (1/RT) from Apple's mobile devices were in line with those found in the literature (see Evans *et al*., 2019, for review). In order to increase validity and standardisation, all instructions were administered to participants in written form for both the morning and afternoon session. This study consisted of two parts. The first part was the mobile Psychomotor

Vigilance Task (m-PVT) reaction time test, which was a modified version of Dinges and Powell's (1985) Psychomotor Vigilance Task. The m-PVT was run on the Dolphin Web Browser mobile application. The second part was the demographic questionnaire that was distributed within Qualtrics Surveys mobile application. In this modified version, the mobile Psychomotor Vigilance Task (m-PVT) (see Figure 72), participants were presented with on-screen instructions and a button at the end that read 'Start'. In each trial, participants were shown a black screen background, and at the centre of the screen they would be presented with a large red fixation circle. The red fixation circle (i.e., inter-stimulus interval) would remain on the screen for a randomised duration that lasted between 2 – 10 seconds, which was then followed by a yellow stimulus counter. As soon as the inter-stimulus interval reached the randomised duration, a yellow stimulus counter appeared counting up in milliseconds from 0 – 5 seconds where it would lapse (i.e., error of omission for 0.5 seconds) and begin the next trial, or until the participant tapped on the screen. Once the participant tapped on the screen, their reaction time (i.e., stimulus) would be displayed for 0.5 seconds. At the end of each trial, a black background would appear on-screen for 0.5 seconds. There were 53 trials in total that lasted approximately 5 minutes. Kribbs and Dinges (1994) found that after a maximum of three trials, the practice effect for the PVT was removed. This study implemented only three practice trials to ensure participants were fully aware of the task, which were removed from final analyses. If participants responded prematurely during any trial (i.e., before the timer commenced counting up), the trial would reset. To also ensure participants were made aware of their premature response, the following message in red was displayed on the centre of the screen, 'You clicked too early! This trial will be reset.' A visual illustration of the mobile Psychomotor Vigilance Task (m-PVT) is presented in Figure 73.

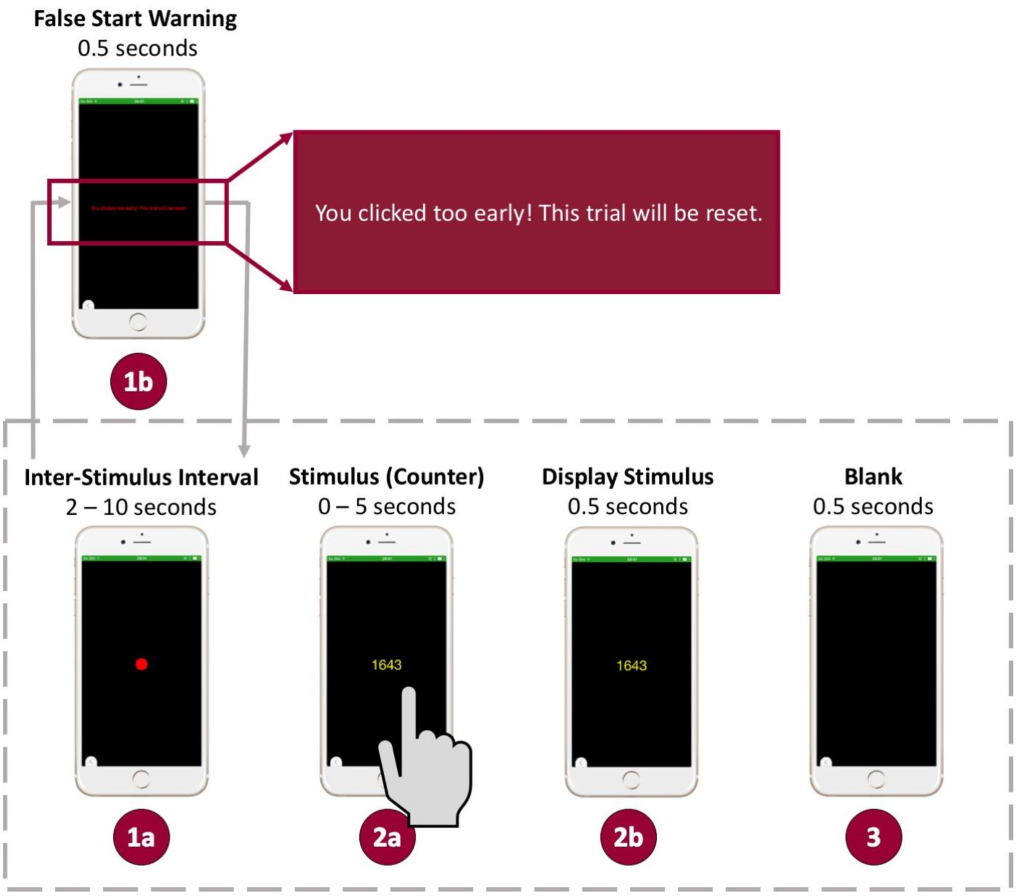


Figure 72: Mobile Psychomotor Vigilance Task (m-PVT) timeline

1a. Participants were presented with a large red circle (i.e., inter-stimulus interval), which appeared for a randomised duration between 2 – 10 seconds.

1b. If participants responded prematurely, a false start warning message appeared informing them that they clicked too early and that the trial would be reset.

2a. As soon as the inter-stimulus interval reached the randomised duration, a yellow stimulus counter appeared counting up in milliseconds from 0 – 5 seconds where it would lapse (i.e., error of omission for 0.5 seconds) and begin the next trial, or until the participant had tapped on the screen.

2b. Once the participants had tapped on the screen, their reaction time (i.e., stimulus) would be displayed for 0.5 seconds.

3. At the end of each trial, a black background would appear on-screen for 0.5 seconds.

Figure 73: Visual illustration of the Mobile Psychomotor Vigilance Task (m-PVT)

6.3.6.1: Simulated workload: Cognitive performance tasks

A series of online cognitive performance tasks were implemented in Qualtrics and administered using the iPhone 6s Plus to simulate high workload. The four counterbalanced cognitive performance tasks used in this study were; the Cambridge Semantic Battery (see Adlam, Patterson, Bozeat & Hodges, 2010, for review), the Reasoning Test (see Baddeley, 1968, for review), the search and memory task (SAM) (see Parkes, 1995, for review), and the Semantic Processing Test (see Baddeley, 1981, for review). A summary of each of these four cognitive performance tasks are presented below.

6.3.6.1.1 Cambridge Semantic Battery

This test was a modified online version of the Cambridge Semantic Battery Camel and Cactus test (CCT), to assess semantic memory, episodic memory and other aspects of cognitive processing (Adlam *et al*., 2010). Research carried out by Rossion and Pourtois (2004) had previously identified that the addition of colour unambiguously improved response time (RT) and naming accuracy. Therefore, the images that were used in this simulated workload study came from Moreno-Martinez and Montoro's (2012) depository, which was in itself an ecological high-quality colour alternative set of pictures from Snodgrass and Vanderwar's (1980) original set of pictures, which have been extensively used in experimental and clinical studies that are based on semantic and episodic memory. For this task, participants were presented with a matching task, whereby there were five pictures, one at the top (i.e., prime) and four at the bottom (i.e., one target and three distractors). Participants were instructed to select one of the four bottom pictures that had the best semantic relationship with the prime at the top (e.g., which goes with camel; sunflower, tree, rose or cactus?) (see Figure 74, for study presentation). Images from 20 categories (see Table 5) were randomly selected from Moreno-Martinez and Montoro's (2012) high-quality colour depository for this study. All distractors used in this study were counterbalanced to ensure that they did not appear in the same location. As a result, any given distractor could only appear up to a maximum of three time. There were 43 randomised trials in total (3 practice trials, 40 test trials). Based on previous pilot studies, this task took on average 5 minutes to complete.

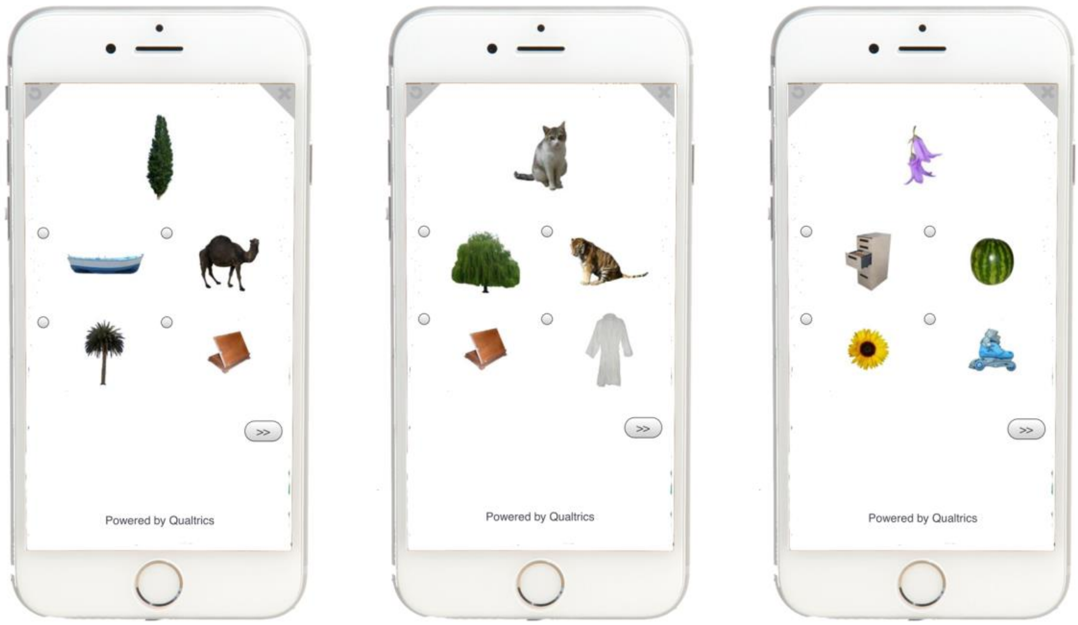


Figure 74: Modified online version of the Cambridge Semantic Battery Camel and Cactus test (CCT)

Living	Non-Living
Animals	Buildings
Birds	Clothing
Body parts	Desk stationary
Flowers	Furniture
Fruits	Kitchen utensils
Insects	Musical instruments
Marine creatures	Sports games
Nuts	Tools
Trees	Vehicles
Vegetables	Weapons

Table 5: Moreno-Martinez and Montoro's (2012) Image Categories

6.3.6.1.2 Reasoning Test

This cognitive test was a modified online version of the syntax reasoning test (see Baddeley, 1968, for review), which is a psycholinguistics high mental processing reasoning task that involves the understanding of sentences that were of various levels of syntactic complexity. This task comprised of 32 questions rather than the original 64 battery of questions. For this task, participants were instructed to select the correct reasoning syntax by determining whether the statement was either true or false based on the order of the letters 'A' and 'B', which were directly followed by the letters AB or BA. For example, 'B is followed by A' (see Figure 75, for study presentation). Unlike Baddeley's (1968) experimental paradigm, there were no time constraints (i.e., 3 minutes) allocated in this version. Instead, this task ended when participants completed all 35 randomised questions (3 practice trials, 32 test trials). Based on previous pilot studies, this task took on average 7 minutes to complete.

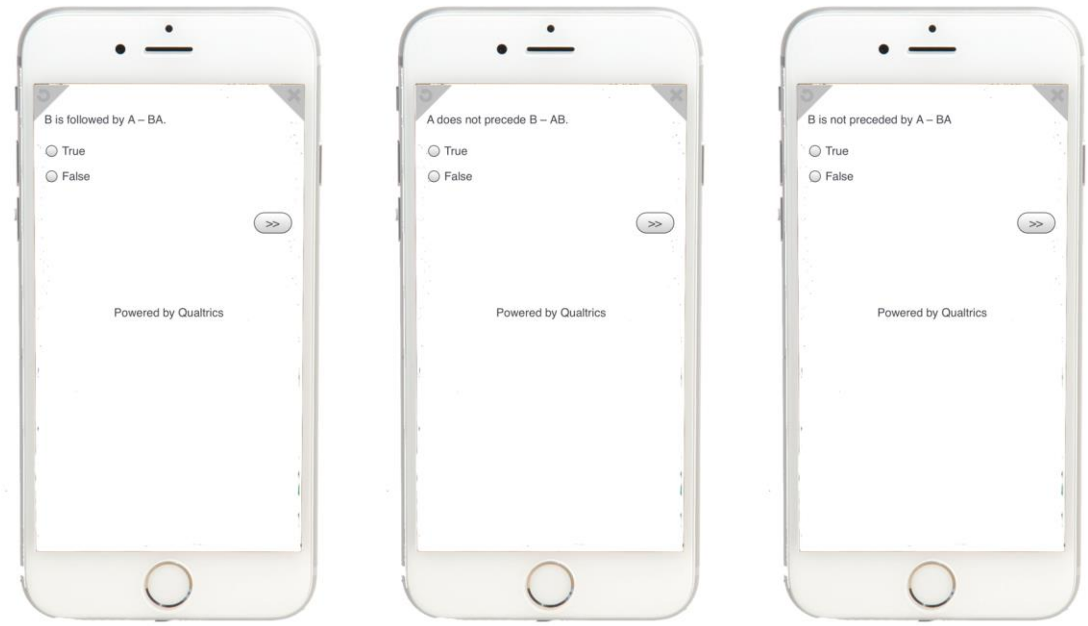


Figure 75: Syntax reasoning test

6.3.6.1.3 Search and Memory task (SAM)

This task was a modified online version of the search and memory task (SAM) (see Parkes, 1995, for review). In this task, participants were presented with a string of 100 randomised letters, which were randomly generated using Matlab® 2014b (The MathWorks Inc., Massachusetts, USA). The 100 letters were presented across five rows (i.e., each row consisted of 20 letters). This additional modification was introduced to ensure that all 100 letters were visible within the iPhone 6s Plus display. Participants were instructed that above the sequence of random letters, they will be shown a single letter. Their task was to determine whether the large letter was either ‘Present’ or “Not Present’ in the sequence of random letters (see Figure 76). The SAM task ended when participants completed all 22 randomised questions (2 practice trials, 20 test trials). Based on previous pilot studies, this task took on average 10 minutes to complete.

Figure 76: Search and Memory (SAM) Task

6.3.6.1.4 Semantic Processing Test

This test was a modified online version of the Semantic Processing Test (see Baddeley, 1981, for review), which was designed to measure the speed of retrieval of general knowledge information. In this test, participants were presented with a sentence and were asked to determine whether said sentence was either false (e.g., onions move around searching for food) or true (e.g., psychiatrists have a profession) (see Figure 77, for study presentation). The Semantic Processing Test ended when participants completed all 53 randomised questions (3 practice trials, 50 test trials). Based on previous pilot studies, this task took on average 5 minutes to complete.

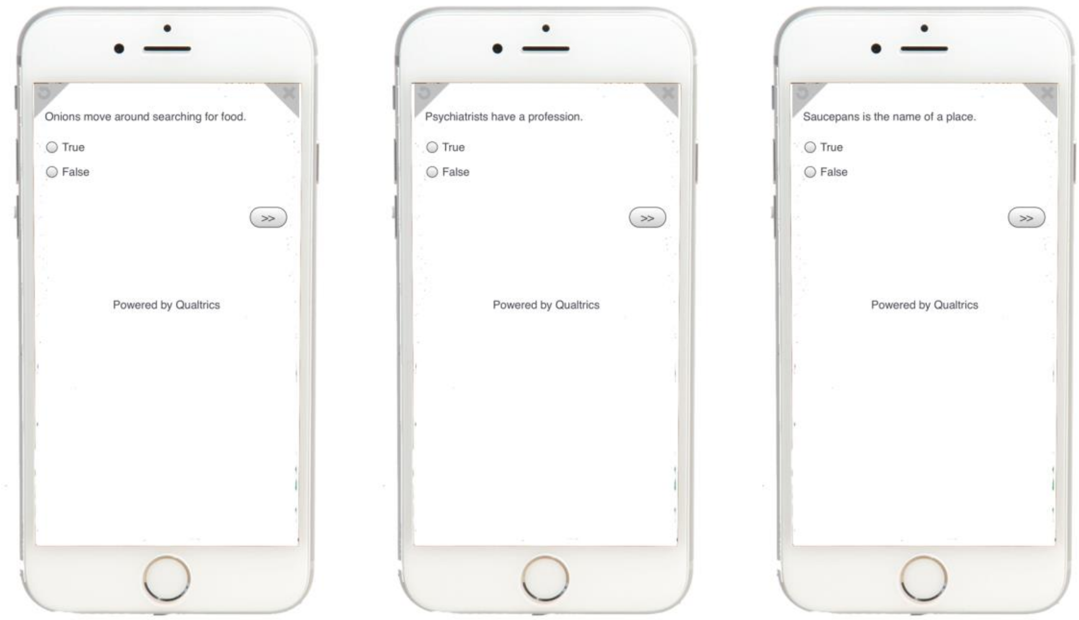


Figure 77: Semantic processing test

6.4: Results

The aim of this study was to investigate whether a shorter mobile version of the online 10-minute mobile Psychomotor Vigilance Task (online 10-min m-PVT) i.e., online 5-min m-PVT could be used to provide an objective indicator of simulated workload fatigue. IBM's Statistical Package for the Social Sciences (SPSS) version 23 for Mac was used to analyse the data. A total of 7,649 test trials (i.e., 50 test trials per session) were submitted for data analyses, with all 458 practice trials (i.e., 3 practice trials per session) excluded from final analyses from all 8,107 recorded trials. It is important to note that all mobile devices running the online 5-min m-PVT were administered through the Dolphin internet browser and were connected using Cardiff University's eduroam Wi-Fi roaming service. Therefore, in rare occasions when the Wi-Fi connectivity dropped, participant's trials were lost and thus not recorded. As a result, a total of 1.95 per cent (n = 161) trials of all potential 8,268 trials (i.e., 468 practice and 7,800 test) were lost and not recorded. In addition, as discussed in the previous study, due to the large variability in the reporting of the PVT outcome metrics (Basner & Dinges, 2011), this study also adopted the three most frequently reported PVT outcome analyses; mean reaction time (RT), mean speed response (1/RT) and mean number of lapses, respectively. Moreover, based on Basner and Dinges (2011) recommendations, all 7,649 test trials with reaction time (RTs) < 100 ms (i.e., false start), which accounted for .04 per cent (n = 3) and RTs ≥ 500 ms (i.e., number of lapses), which accounted for 7.48 per cent (n = 572), were considered for exclusion from the final mean reaction time (RT) and mean speed response (1/RT) analyses. All 7.48 per cent (n = 572) of RTs ≥ 500 ms (i.e., number of lapses) were analysed separately.

6.4.1: Reliability and validity

The following subsections address the reliability and validly of the online 5-min m-PVT by highlighting the importance of; reporting the effect size to provide a confidence level in the reported *p*-values, reporting the intraclass correlation coefficient (ICC), which is a widely used reliability analysis, as well as reporting Mauchly's Test of Sphericity, which tests the hypothesis that the variances of the differences between conditions are equal.

6.4.1.1: Effect size

Chapter 5 outlines the rationale for reporting effect size. Therefore, the following partial eta-squared (η_p^2) threshold criteria were also used in this study; .01 small, .06 medium, and .14 large (Cohen, 1973; Cohen, 1988).

6.4.1.2: Reliability

Chapter 5 outlines the rationale for reporting the intraclass correlation coefficient (ICC). The ICC estimates and their 95% confidence intervals based on a mean-rating (k = 20), absolute-agreement, 2-way mixed-effects model revealed that there was an excellent degree of reliability (see Ko & Li, 2016, for review) between all 20 mean reaction times (RTs) i.e., the low pre-simulated workload online 5-min m-PVT (1-minute, 2 minutes, 3 minutes, 4 minutes, and 5 minutes), the low post-simulated workload online 5-min m-PVT (1-minute, 2 minutes, 3 minutes, 4 minutes, and 5 minutes), the high pre-simulated workload online 5-min m-PVT (1-minute, 2 minutes, 3 minutes, 4 minutes, and 5 minutes), and the high post-simulated workload online 5-min m-PVT (1-minute, 2 minutes, 3 minutes, 4 minutes, and 5 minutes). The average measure ICC was .970 with a 95% confidence interval from .955 to .982, $F(38, 722) = 35.56$, $p < .001$.

6.4.2: Mauchly's Test of Sphericity for mean reaction time (RT)

The assumption of sphericity was examined using the Mauchly's Test of Sphericity. For the mean reaction time (RT) analyses, all Mauchly's Test of Sphericity indicated that the assumption of sphericity had not been violated: time on task, $\chi^2(9) = 10.16$, $p = .338$; simulated workload × time on task, $\chi^2(9) = 10.31$, $p = .326$; time of day × time on task, $\chi^2(9) = 4.80$, $p = .852$; and simulated workload × time of day × time on task, $\chi^2(9) = 7.26$, $p = .611$. In addition, the Greenhouse–Geisser test was reported for simulated workload, time of day, and simulated workload × time of day, as nothing was known about sphericity.

6.4.3: Mean reaction time (RT)

Figure 78 and Figure 79 presents the illustrated mean reaction times (RTs) across the different conditions split into high simulated workload and low simulated workload. Mean RTs were submitted to a 2 × 2 × 5 three-way repeated analysis of variance (ANOVA) comparing 2 × simulated workload (high simulated workload i.e., completing a 30-minutes battery of cognitive performance tasks or low simulated workload i.e., watching a television show) × 2 time of day (Time 1: pre-simulated workload online 5-min m-PVT, or Time 2: post-simulated workload online 5-min m-PVT) × 5 time on task (1-minute; 2 minutes; 3 minutes, 4 minutes; or 5 minutes).

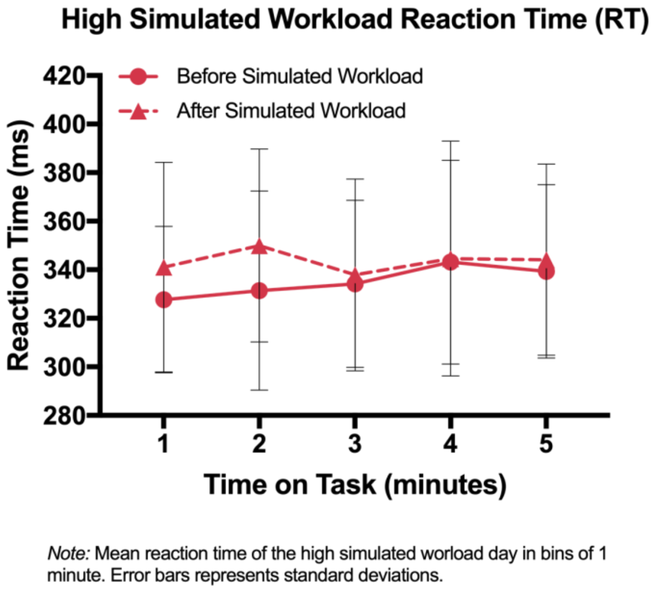


Figure 78: Mean reaction times (RTs) for high simulated workload and time on task

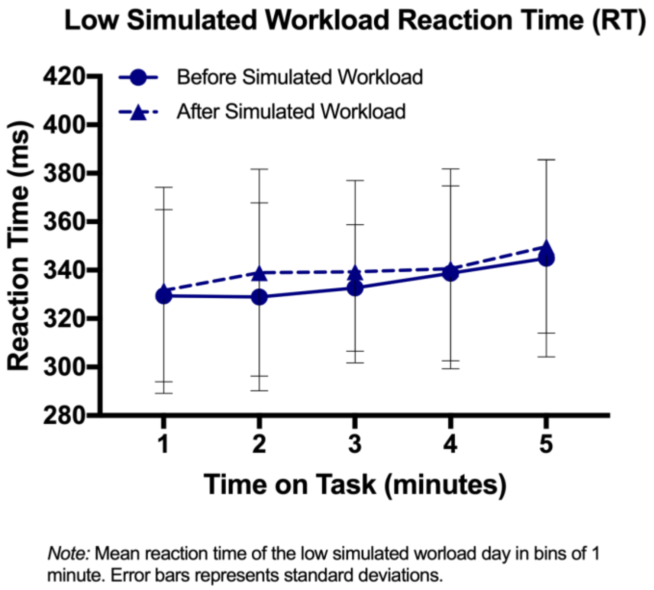


Figure 79: Mean reaction times (RTs) for low simulated workload and time on task

There was a significant main effect of time of day, Greenhouse–Geisser = 1.00, $F(1, 38)$, 7.11, $p = .011$, $\eta_p^2 = .16$, indicating a large effect size (Cohen, 1973; Cohen, 1988; Levine & Hullett, 2002), which provides a high level of confidence in the reported *p*-value. Furthermore, there was also a significant main effect of time on task, Sphericity Assumed, $F(4, 152)$, 7.30, $p < .001$, $\eta_p^2 = .16$, indicating a large effect size (Cohen, 1973; Cohen, 1988; Levine & Hullett, 2002), which also provides a high level of confidence in the reported p-value. All other main effect interactions were not significant. There was no significant: main effect of simulated workload, Greenhouse–Geisser = 1.00, $F(1, 38)$, .496, $p = .486$, $\eta_p^2 = .01$; two-way interaction, simulated workload × time of day, Greenhouse–Geisser = 1.00, $F(1, 38)$, .649, $p = .426$, $\eta_p^2 = .02$; two-way interaction, simulated workload × time on task, Sphericity Assumed, $F(4, 152)$, 2.29, $p = .062$, $\eta_p^2 = .06$; two-way interaction, time of day × time on task, Sphericity Assumed, $F(4, 152)$, 1.81, $p = .128$, $\eta_p^2 = .05$; and three-way interaction, simulated workload × time of day × time on task, Sphericity Assumed, $F(4, 152)$, .904, $p = .463$, $\eta_p^2 = .02$.

6.4.3.1: Mean reaction time (RT) Post-Hoc tests

The main effect of time of day for mean reaction times (RTs) is illustrated in Figure 80 and was followed by post-hoc tests with Bonferroni correction for multiple comparisons. Post-hoc tests also showed that participants mean RTs were also significantly faster in the before simulated workload session ($M = 335.04$ ms, $SE = 4.79$ ms) when compared to mean RTs in the after simulated workload session ($M = 341.81$ ms, $SE = 5.54$ ms) $p = .011$.

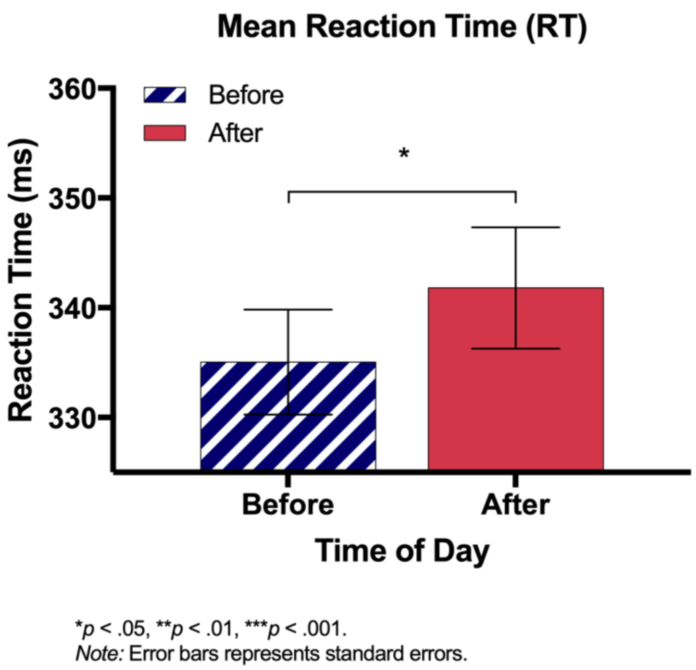
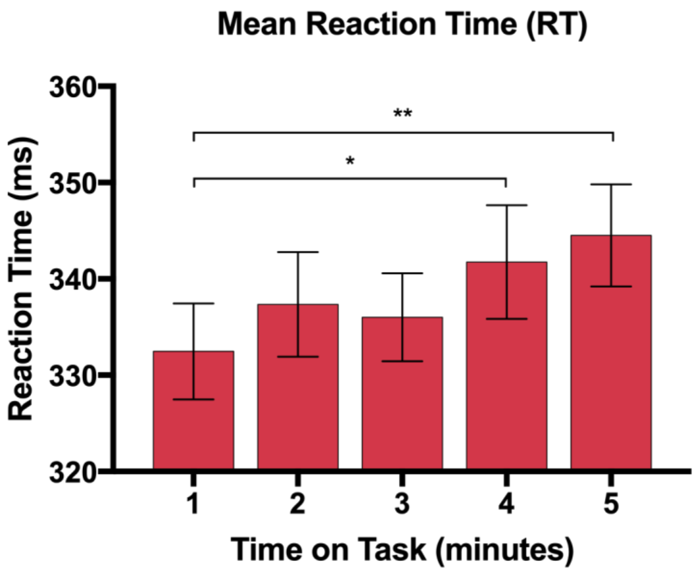


Figure 80: Mean reaction times (RTs) for time of day (before simulated workload or after simulated workload)

The main effect of time on task for the mean reaction times (RTs) is illustrated in Figure 81 and was followed by post-hoc tests with Bonferroni correction for multiple comparisons. Post-hoc *t*-tests showed that participants mean RTs were significantly slower after 4 minutes (M = 341.75 ms, SE = 5.91 ms) when compared to mean RTs after 1-minute (M = 332.47 ms, SE = 4.98 ms) p = .019, on the online 5-min m-PVT temporal attention task. As expected, participants mean RTs were also significantly slower after 5 minutes (M = 344.52 ms, SE = 5.31 ms) when compared to mean RTs after 1-minute (M = 332.47 ms, SE = 4.98 ms) p = .001. There was no significant difference when comparing mean RTs after 1-minute (M = 332.47 ms, SE = 4.98 ms) and 2-minutes (M = 337.35 ms, SE = 5.44 ms) p = .212, as well as when comparing mean RTs after 1-minute (M = 332.47 ms, SE = 4.98 ms) and 3-minutes (M = 336.03 ms, SE = 4.57 ms) p = 1.00.



*p < .05, **p < .01, ***p < .001.
Note: Mean RTs are presented in bins of 1-minute intervals. Error bars represents standard errors.

Figure 81: Mean reaction times (RTs) for time on task (1-minute intervals)

6.4.4: Mauchly's Test of Sphericity for mean speed response (1/RT)

The assumption of sphericity was examined using the Mauchly's Test of Sphericity. For the mean speed response (1/RT) analyses, all Mauchly's Test of Sphericity indicated that the assumption of sphericity had not been violated: time on task, $\chi^2(9) = 8.92$, $p = .445$; simulated workload × time on task, $\chi^2(9) = 11.53$, $p = .242$; time of day × time on task, $\chi^2(9) = 3.95$, $p = .915$; and simulated workload × time of day × time on task, $\chi^2(9) = 10.49$, $p = .313$. In addition, Greenhouse–Geisser test was reported for; simulated workload, time of day, and simulated workload × time of day, as nothing was known about sphericity.

6.4.5: Mean speed response (1/RT)

Figure 82 and Figure 83 present the illustrated mean speed responses (1/RTs) across the different conditions split into high simulated workload and low simulated workload. 1/RTs were submitted to a 2 × 2 × 5 three-way repeated analysis of variance (ANOVA) comparing 2 × simulated workload (high simulated workload i.e., completing a 30-minutes battery of cognitive performance tasks or low simulated workload i.e., watching a television show) × 2 time of day (Time 1: pre-simulated workload online 5-min m-PVT, or Time 2: post-simulated workload online 5-min m-PVT) × 5 time on task (1-minute; 2 minutes; 3 minutes, 4 minutes; or 5 minutes).

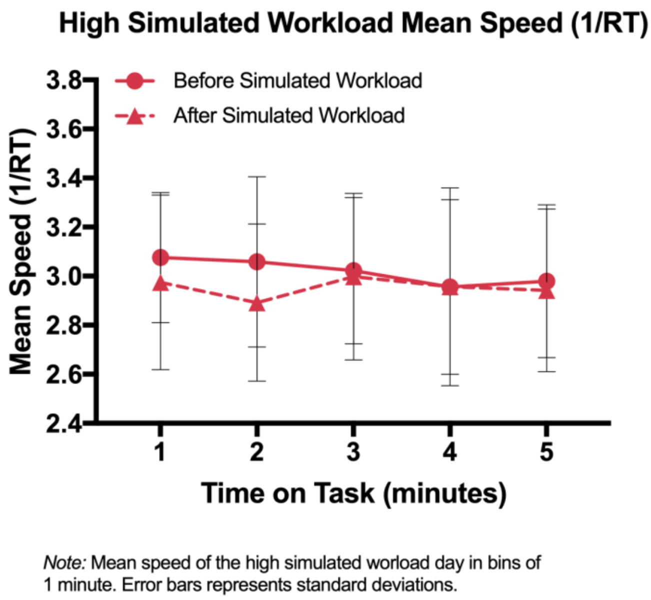
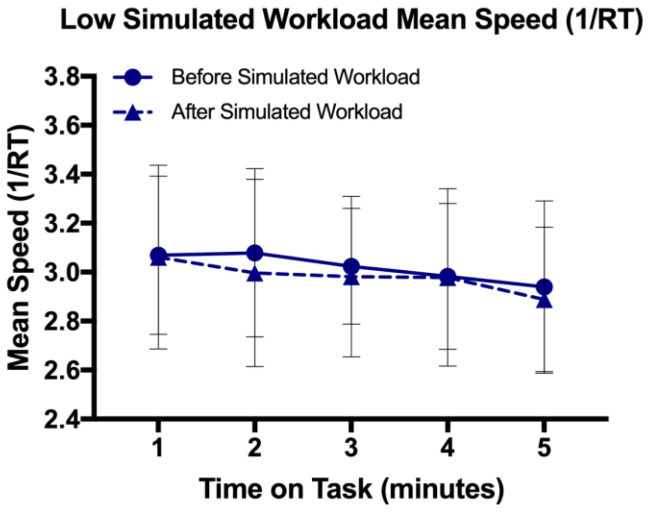


Figure 82: Mean speed responses (1/RTs) for high simulated workload and time on task



Note: Mean speed of the low simulated worload day in bins of 1 minute. Error bars represents standard deviations.

Figure 83: Mean speed responses (1/RTs) for low simulated workload and time on task

There was no significant main effect of the simulated workload, Greenhouse–Geisser = 1.00, $F(1, 38)$, .371, $p = .546$, $\eta_p^2 = .01$. There was a significant main effect of time of day, Greenhouse–Geisser = 1.00, $F(1, 38)$, 5.48, $p = .025$, $\eta_p^2 = .13$, indicating a medium effect size (Cohen, 1973; Cohen, 1988; Levine & Hullett, 2002), which provides a moderate level of confidence in the reported *p*-value. Furthermore, there was also a significant main effect of time on task, Sphericity Assumed, $F(4, 152)$, 7.43, $p < .001$, $\eta_p^2 = .16$, indicating a large effect size (Cohen, 1973; Cohen, 1988; Levine & Hullett, 2002), which provides a high level of confidence in the reported *p*-value. All other main effect and interactions were not significant. There was no significant: main effect of the simulated workload, Sphericity Assumed, $F(1, 38)$, .371, $p = .546$, $\eta_p^2 = .01$; two-way interaction, simulated workload × time of day, Greenhouse–Geisser = 1.00, $F(1, 38)$, .552, $p = .462$, $\eta_p^2 = .01$; two-way interaction between simulated workload × time on task, Sphericity Assumed, $F(4, 152)$, 2.17, $p = .075$, $\eta_p^2 = .05$; two-way interaction, time of day × time on task, Sphericity Assumed, $F(4, 152)$, 2.15, $p = .077$, $\eta_p^2 = .05$; and three-way interaction, simulated workload × time of day × time on task, Sphericity Assumed, $F(4, 152)$, .942, $p = .441$, $\eta_p^2 = .02$.

6.4.5.1: Mean speed response (1/RT) Post-Hoc tests

The main effect of time of day for mean speed responses (1/RTs) is illustrated in Figure 84 and was followed by post-hoc tests with Bonferroni correction for multiple comparisons. Post-hoc tests showed that participants mean 1/RTs were significantly better in the before simulated workload session (M = 3.02, SE = .04) when compared to mean 1/RTs in the after simulated workload session (M = 2.97, SE = .05) p = .025.

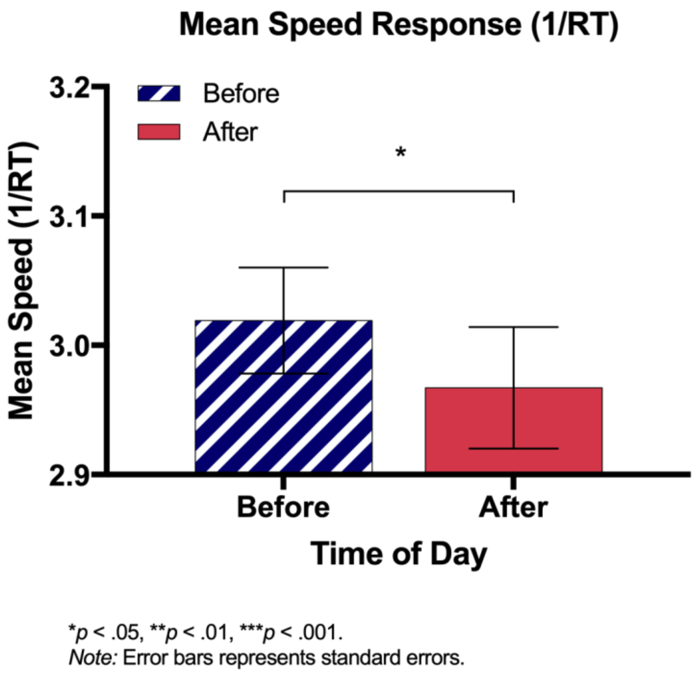


Figure 84: Mean speed responses (1/RTs) and for time of day (before simulated workload or after simulated workload)

The main effect of time on task for the mean speed responses (1/RTs) is illustrated in Figure 85 and was followed by post-hoc t-tests with Bonferroni correction for multiple comparisons. Post-hoc tests showed that participants mean 1/RTs were significantly slower in their 1/RTs after 4 minutes (M = 2.97, SE = .05) when compared to mean 1/RTs after 1-minute (M = 3.05, SE = .04) p = .034, on the m-PVT temporal attention task. As expected, participants were also significantly slower in their mean 1/RTs after 5 minutes (M = 2.94, SE = .04) when compared to mean 1/RTs after 1-minute (M = 3.05, SE = .04) p < .001, on the m-PVT temporal attention task. There was no significant difference when comparing mean 1/RTs from the first minute (M = 3.05, SE = .04) and second minute (M = 3.01, SE = .05) p = .362, as well as when comparing the first minute (M = 3.05, SE = .04) and third minute (M = 3.01, SE = .04) p = .585.

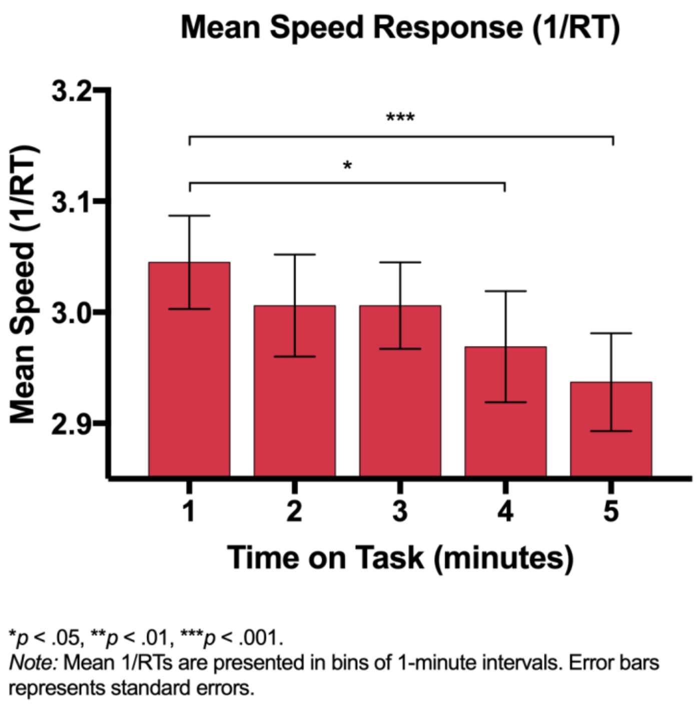


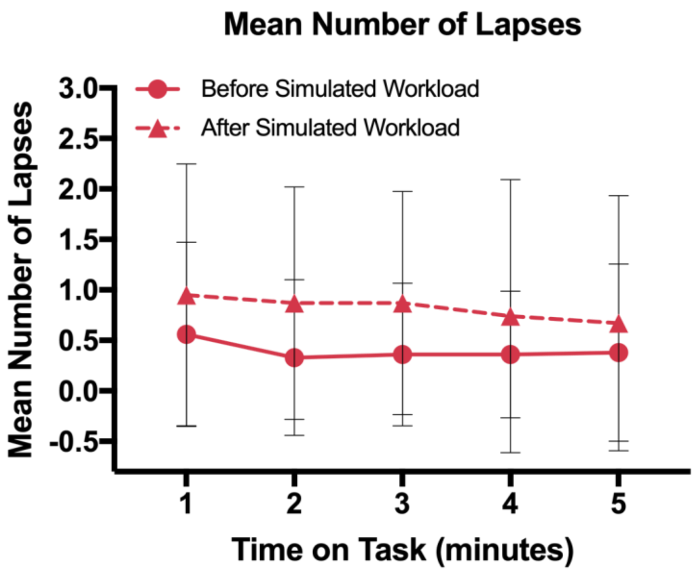
Figure 85: Mean speed responses (1/RTs) for time on task (1-minute intervals)

6.4.6: Mauchly's Test of Sphericity for mean number of lapses

The assumption of sphericity was examined using the Mauchly's Test of Sphericity. For the mean number of lapses analyses, Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated for the three-way interaction simulated workload × time of day × time on task, $\chi^2(9) = 21.39$, p = .011. Therefore, the Huynh-Feldt test were reported instead of Sphericity Assumed since Greenhouse–Geisser Epsilon was greater than .75 for the time on task, and time of day × time on task. All other Mauchly's Test of Sphericity indicated that the assumption of sphericity had not been violated: time on task, $\chi^2(9) = 6.23$, p = .717; simulated workload × time on task, $\chi^2(9)$ = 4.70, p = .867; and time of day × time on task, $\chi^2(9) = 13.58$, p = .139. In addition, Greenhouse–Geisser test was reported for; simulated workload, time of day, and simulated workload × time of day, as nothing was known about sphericity.

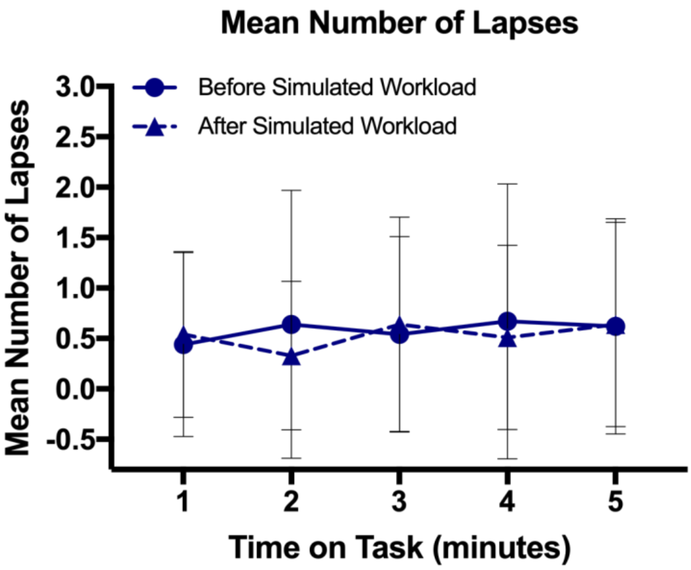
6.4.7: Mean number of lapses

From all test trials, a total of 7.48 per cent (n = 572) of RTs ≥ 500 ms (i.e., number of lapses) were submitted for data analyses. Figure 86 and Figure 87 present the illustrated mean number of lapses across the different conditions split into high simulated workload and low simulated workload. The mean number of lapses were submitted to a 2 × 2 × 5 three-way repeated analysis of variance (ANOVA) comparing 2 × simulated workload × 2 time of day × 5 time on task (1-minute; 2 minutes; 3 minutes, 4 minutes; or 5 minutes).



Note: Mean number of lapses of the high simulated worload day in bins of 1 minute. Error bars represents standard deviations.

Figure 86: Mean number of lapses for time on task for high simulated workload



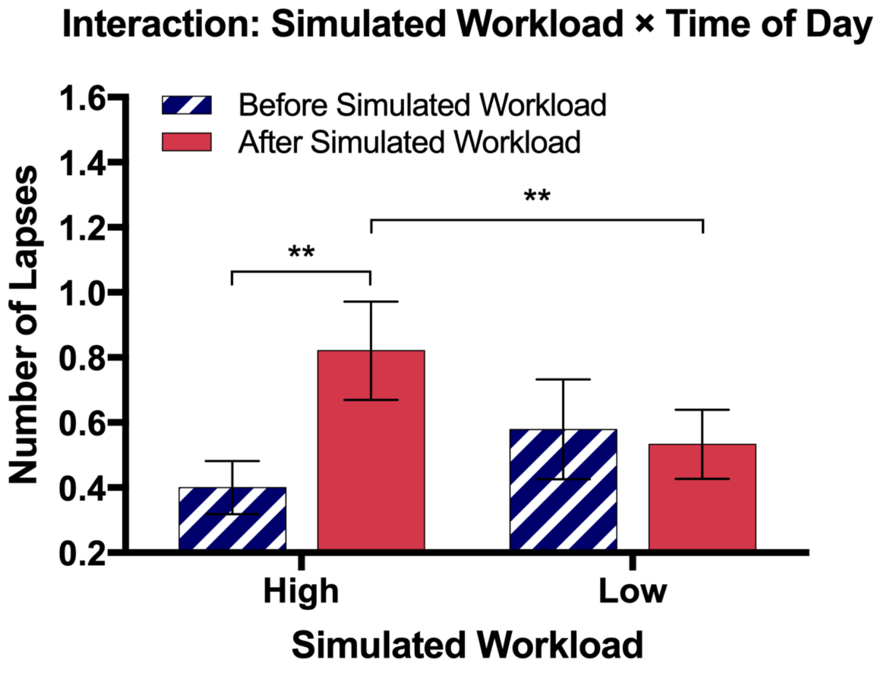
Note: Mean number of lapses of the low simulated worload day in bins of 1 minute. Error bars represents standard deviations.

Figure 87: Mean number of lapses for time on task for low simulated workload

There was a significant main effect of time of day, Greenhouse–Geisser = 1.00, $F(1, 38)$, 6.69, $p = .014$, $\eta_p^2 = .15$, indicating a large effect size (Cohen, 1973; Cohen, 1988; Levine & Hullett, 2002) as well as a significant interaction between simulated workload × time of day, Greenhouse–Geisser = 1.00, $F(1, 38)$, 10.28, $p = .003$, $\eta_p^2 = .21$, indicating a large effect size (Cohen, 1973; Cohen, 1988; Levine & Hullett, 2002), which both provide a high level of confidence in the reported *p*-value.. All other main effects and interactions were not significant. The main effect of simulated workload, Greenhouse–Geisser = 1.00, $F(1, 38)$, .649, $p = .425$, $\eta_p^2 = .02$; the main effect of time on task, Sphericity Assumed, $F(4, 152)$, .265, $p = .900$, $\eta_p^2 = .01$; two-way interaction, simulated workload × time on task, Sphericity Assumed, $F(4, 152)$, 1.29, $p = .278$, $\eta_p^2 = .03$; two-way interaction, time of day × time on task, Sphericity Assumed = 1.00, $F(4, 152)$, .568, $p = .686$, $\eta_p^2 = .02$; and three-way interaction, simulated workload × time of day × time on task, Huynh-Feldt = .845, $F(3.38, 128.37)$, .953, $p = .425$, $\eta_p^2 = .02$.

6.4.7.1: Mean number of lapses Post-Hoc tests

The interaction between simulated workload × time of day is illustrated in Figure 88 and was followed by post-hoc *t*-tests with Bonferroni correction for multiple comparisons. There was also no significant difference in the mean number of lapses before the simulated workload condition for both the high simulated workload (M = .40, SE = .08) and the low simulated workload (M = .58, SE = .15), p = .078, which seems to indicate that the mean number of lapses between both high and low simulated workload conditions did not differ significantly. However, there was a significant difference in the mean number of lapses after the simulated workload condition for both the high simulated workload (M = .82, SE = .15) and the low simulated workload (M = .53, SE = .11), p = .006. These findings seem to indicate that participants generated a greater mean number of lapses after completing 30-minutes of cognitive performance tasks (i.e., high simulated workload condition) when compared to after watching The Big Bang Theory (i.e., low simulated workload condition). Moreover, it was also found that on the high simulated workload condition, participants had a significantly higher mean number of lapses after completing the 30-minutes of cognitive performance tasks (M = .82, SE = .15) when compared to before starting the 30-minutes of cognitive performance tasks (M = .40, SE = .08), p = .001. In contrast, there was no significant difference in the mean number of lapses for the low simulated workload condition before (M = .58, SE = .15) watching a television show (i.e., The Big Bang Theory) and after watching The Big Bang Theory (M = .53, SE = .11), p = .604. These findings seem to indicate that perhaps simulated workload may elicit sensitivity levels of fatigue.



*p < .05, **p < .01, ***p < .001.
Note: Error bars represents standard errors.

Figure 88: Interaction: Simulated workload × time of day

6.5: Discussion

The aim of this study was to investigate whether a shorter mobile version of the online 10-minute mobile Psychomotor Vigilance Task (online 10-min m-PVT) i.e., online 5-min m-PVT could be used to provide an objective indicator of simulated workload fatigue. Firstly, intraclass correlation coefficient (ICC) analysis revealed an excellent degree of reliability in this study (see Ko & Li, 2016, for review) across all 20 mean reaction Times (RTs) which indicates that the online 5-min m-PVT was able to generate highly reliable RTs across all participants. Secondly, these findings also have a higher level of confidence as all significant *p*-values comprised of either a medium or high partial eta-squared (η_p^2) effect size.

According to Basner and Dinges (2011) the most commonly reported PVT outcome was number of lapses. An examination of the number of lapses (i.e., reaction times ≥ 500 ms) revealed that for the high simulated workload condition, participants' generated a significantly greater number of lapses after completing a series of online cognitive tasks i.e., Cambridge semantic battery (Adlam *et al*., 2010), the reasoning test (Baddeley, 1968), the search and memory task (SAM) (Parkes, 1995), and the semantic processing test (Baddeley, 1981) designed to elicit mental workload when compared to the low simulated workload condition i.e., after watching one episode of The Big Bang Theory. These findings support Evirgen *et al.*'s (2015) study who found that participants who were assigned to a high workload condition had significantly slower PVT reaction times after completing 60-minutes of continuous cognitive tasks when compared to PVT reaction times before the cognitive tasks. In contrast, no significant differences were found in participants PVT reaction times before and after completing a 40-minute resting period i.e., low workload condition. However, the present study also found that participants generated a greater mean number of lapses after completing a series of online cognitive tasks when compared to after watching one episode of The Big Bang Theory. These findings seem to further support the idea that perhaps high simulated workload may elicit sensitivity levels of fatigue. These findings are consistent with Gui *et al*. (2015) who have identified that after continuous and prolonged mental workload, individuals will begin to show reduced behavioural performance. This is further reinforced by Warm *et al*. (2008) who have pointed out that the PVT consumes significant mental resources and as a result, vigilance tasks cannot immediately replenish consumed mental resources.

In this study it was found that participants had significantly slower reaction times (RTs) and greater speed responses (1/RTs), as well as generated more number of lapses after the simulated workload condition (i.e., either after completing a series of online cognitive tasks, or watching one episode of The Big Bang Theory) when compared to before completing the simulated workload. These results support Zammouri *et al*. (2018) findings who have identified that after continuous and prolonged mental workload, individuals will begin to show reduced behavioural performance. As a result, tasks that require continuous vigilance e.g., completing a series of online cognitive tasks consumes significantly more mental resources (Warm *et al.*, 2008). This is further supported by Eggemeier *et al.* (1991) who outlined that in order to be able to meet task demands, there is usually a required amount of operator resources needed, for a person to complete the task over a given period of time (Longo, 2015; Longo, 2017).

Findings from this study also support previous research that have identified an increase in fatigue results in impaired alertness (Dorrian *et al.*, 2000; Gillberg *et al.*, 1994), whereby time on task (i.e., sustained attention), as measured in mean RTs and mean 1/RTs significantly differed after 4 minutes of continuous performance using the online 5-min m-PVT. Therefore, these findings were consistent to those of the previous study, which also found that RTs and 1/RTs significantly differed after 4 minutes of continuous performance (Evans *et al.*, 2019). As a result, these findings further support previous work, which suggested that sustained attention drops with prolonged duration of task (Dinges & Powell, 1988; Dinges & Powell, 1989; Doran *et al.*, 2001). These findings seem to further validate that the online 5-min m-PVT is sensitive enough to detect levels of fatigue, even after only 4 minutes on the task. These findings support the time-on-task effect (Langner & Eickhoff, 2013; Mackworth, 1948; Mackworth, 1968). However, the interaction between simulated workload and time on task revealed that both the RT and 1/RT outcome metrics were only significant for the low simulated workload condition after 5 minutes, with no significant differences found in both the RT and 1/RT outcome metrics for the high simulated workload condition. However, inter-individual variation could explain these inconsistencies. Lim *et al.* (2010) found substantial inter-individual variation in participant's RTs using the PVT by the end of the task. Therefore, the time on task effect findings in this study seem to indicate that caution needs to be taken going forward as the shorter version of the 'gold standard' 10-min PVT i.e., online 5-min m-PVT may not be long enough to detect levels of fatigue. These findings do not support previous studies which have been able to demonstrate that shorter versions of the 'gold standard' 10-min PVT can exhibit the time-on-task effect after 5-min PVT (Roach *et al*., 2006), and even after only 3-min PVT (PVT-B) (Basner *et al*., 2011). Therefore, even though in a controlled laboratory environment a shorter online 5-min m-PVT may be sensitive enough to detect levels of fatigue based on simulated workload, this may not be guaranteed in applied frontline safety critical settings. Furthermore, it is important to acknowledge that this study recruited undergraduate students who firstly do not have the type of work pressures that may be expected from frontline safety critical workers. According to Zoer *et al.* (2011), work pressure was the most significant risk factor of railway employees for expressing mental health complaints. In addition, in this study participants were instructed not to consume caffeine (e.g., coffee, energy drinks, tea, etc.) and alcohol during the 24 hours before the study. Such restrictions would not be possible or realistic in frontline safety critical settings. There were additional limitations with the present study which suffered from the same issue as outlined in the previous chapter.

Firstly, the series of online cognitive performance tasks, which were implemented and administered through the online Qualtrics Surveys were solely used for the purpose of eliciting the time-on-task effect (Gui *et al*., 2015; Gunzelmann, *et al*., 2011; Langner & Eickhoff, 2013; Mackworth, 1948; Mackworth, 1968; Satterfield *et al*., 2017; Warm *et al*., 2008) as well as simulating high workload (see Evirgen *et al.*, 2015, for review). Therefore, no further analyses were carried out on these four cognitive performance tasks, as previous pilot studies had identified that generated reaction times (RTs) from these tasks were erratic. As a result, by enabling and integrating the more advanced functions within Qualtrics' JavaScript programming language, additional processing and computational time were introduced (Barnhoorn *et al*., 2015), which may have resulted in erratic RTs ranged between 3 – 20 seconds. Secondly, despite studies outlining the availability to run online experiments as previously pointed out in the earlier study

(Barnhoorn *et al*., 2015; Crump *et al*., 2013; Reimers & Stewart, 2016) and other studies demonstrating comparability between lab based and online based experiments (e.g., de Leeuw & Motz, 2016; Reimers & Stewart, 2007; Schubert *et al*., 2013; Simcox & Fiez, 2014), in rare occasions when the Wi-Fi connectivity dropped, participant's trials were lost and thus not recorded. As a result, a total of 1.95 per cent (n = 161) were never recorded, which despite representing a large reduction of 83.49 per cent Wi-Fi connectivity drop when compared to the previous study (n = 975), are still relatively high and could potentially impact the validity of the m-PVT. Therefore, once again as highlighted in the previous chapter, there is a clear need for the development of an offline iOS mobile app version of the 10-min m-PVT. Further research is now needed to investigate whether an offline iOS mobile app version of the online 10-min m-PVT i.e., offline 10-min m-PVT could also be used to detect sensitivity levels of fatigue in frontline safety critical workers.

6.6: Chapter summary

The aim of this study was to investigate whether a shorter mobile version of the online 10-minute mobile Psychomotor Vigilance Task (online 10-min m-PVT) i.e., online 5-min m-PVT could be used to provide an objective indicator of simulated workload fatigue. 39 (28 female, M = 19.95, SD = 2.15) participants were voluntarily recruited from Cardiff University via the Experimental Management System (EMS) to take part in this study. The study involved participants attending two counterbalanced morning (i.e., 07:30am) sessions; a high workload session (i.e., participants completed a 30-min battery of cognitive performance tasks implemented on a mobile phone) and a low workload session (i.e., participants watched one episode of the television show - The Big Bang Theory). The online 5-min m-PVT was carried out both before (i.e., time 1) and after (i.e., time 2) the simulated workload condition. In this study it was found that for the high simulated workload condition, participants' mean number of lapses were significantly higher after completing a series of online cognitive tasks designed to elicit mental workload (i.e., high simulated workload) when compared to before completing a series of online cognitive tasks. In contrast, there was no significant difference in participants' mean number of lapses before and after watching one episode of The Big Bang Theory (i.e., low simulated workload). However, the interaction between simulated workload and time on task revealed that both the RT and 1/RT outcome metrics were only significant for the low simulated workload condition after 5 minutes, with no significant differences found in both the RT and 1/RT outcome metrics for the high simulated workload condition. Therefore, the time on task effect in this study seem to indicate that caution needs to be taken when administering a shorter version of the 'gold standard' 10-min PVT i.e., online 5-min m-PVT as it may not be sensitive or reliable enough to detect levels of fatigue. However, despite studies outlining the availability to run online experiments (Barnhoorn *et al*., 2015; Crump *et al*., 2013; Reimers & Stewart, 2016) and other studies demonstrating comparability between lab based and online based experiments (e.g., de Leeuw & Motz, 2016; Reimers & Stewart, 2007; Schubert *et al*., 2013; Simcox & Fiez, 2014), in rare occasions when the Wi-Fi connectivity dropped, participant's trials were lost and thus not recorded. As a result, a total of 1.95 per cent (n = 161) were never recorded, which despite representing a large reduction of 83.49 per cent Wi-Fi connectivity drop when compared to the previous study (n = 975), are still relatively high and could potentially impact the validity of the m-PVT. Therefore, once again as highlighted in the previous chapter, there is a clear need for the development of an offline iOS mobile app version of the 10-min m-PVT. Further research is now needed to investigate whether an offline iOS mobile app version of the online 10-min m-PVT i.e., offline 10-min m-PVT could also be used to detect sensitivity levels of fatigue in frontline safety critical workers.

Chapter 7: Developing and validating an offline iOS mobile app version of the online 10-min m-PVT i.e., offline 10-min m-PVT for frontline safety critical workers

7.1: Overview of chapter

The previous chapter identified that in the high simulated workload condition, participants' mean number of lapses were significantly higher after completing a series of online cognitive tasks designed to elicit mental workload (i.e., high simulated workload) when compared to before completing a series of online cognitive tasks. In contrast, there was no significant difference in participants' mean number of lapses before and after watching one episode of The Big Bang Theory (i.e., low simulated workload). Therefore, these findings seem to indicate that the online 5-min m-PVT could be used to provide an objective indicator of simulated workload fatigue. However, there was also an interaction between simulated workload and time on task, which revealed that both the RT and 1/RT outcome metrics were only significant for the low simulated workload condition after 5 minutes, with no significant differences found in both the RT and 1/RT outcome metrics for the high simulated workload condition. Therefore, these findings seem to indicate that caution needs to be taken when administering a shorter offline version of the 'gold standard' 10-min PVT as it may not be sensitive or reliable enough to detect levels of fatigue in frontline safety critical workers such as train drivers, hospital staff, emergency services, law enforcers, etc. The aim of the present chapter is to develop and validate an alternative offline iOS mobile app version of the online 10-min m-PVT i.e., offline 10-min m-PVT to detect levels of fatigue for frontline safety critical workers.

7.2: Introduction and rationale

Previous chapters have outlined, described, and demonstrated that the Psychomotor Vigilance Task (PVT) is highly reliable for assessing cognitive neurobehavioral performance on sustained attention (e.g., Belenky *et al.*, 2003; Dinges *et al.*, 1997; Jewett *et al*., 1999; Lamond *et al*., 2003). As a result, the PVT has been widely administered in numerous studies to be an effective objective cognitive performance measure, across multiple paradigms; the time-on-task effect (e.g., Basner *et al*., 2011; Evans *et al*., 2019; Gui *et al*., 2015; Langner & Eickhoff, 2013; Lim *et al*., 2010; Loh *et al*., 2004; Mackworth, 1948; Mackworth, 1968; Roach *et al*., 2006; Schmidt & van Dongen, 2017; Warm *et al*., 2008), the time-of-day effect (e.g., Carrier & Monk, 2000; Lafrance & Dumont, 2000; Lenne *et al*., 1997; Patkai, 1971; Smith, 1992), and high/low workload conditions (e.g., Arnedt *et al*., 2005; Evirgen *et al*., 2015; Goel *et al*., 2014; Landrigan *et al*., 2004). In addition, the PVT has also been substantially utilised to investigate; sleep quality (de Godoy *et al.,* 2016; Kubo *et al.,* 2016; van Ryswyk *et al.,* 2017), fatigue (Baulk, Biggs, Reid, van den Heuvel, & Dawson, 2008; Gander *et al.,* 2013; Gander *et al.,* 2014a; Gander *et al.,* 2014b; Gander *et al.,* 2015; Kubo *et al.,* 2016; Liu *et al.,* 2016; Roach *et al.,* 2012; Volker, Kirchner, & Bock, 2016), computational modelling of fatigue (Walsh, Gunzelmann, & van Dongen, 2017), performance (Gui *et al.,* 2015; van Ryswyk *et al.,* 2017), well-being (van Ryswyk *et al.,* 2017), fatigue in road drivers (Baulk *et al.,* 2008; Kosmadopoulos *et al.,* 2017; Sparrow *et al.,* 2016), fatigue in aviation pilots (Gander *et al.,* 2013; Gander *et al.,* 2014b; Gander *et al.,* 2015; Gander *et al.,* 2016; Honn *et al.,* 2016; Roach, Petrilli, Dawson, & Lamond, 2012; Signal *et al.,* 2014), fatigue in aviation cabin crew (van den Berg et al., 2015), fatigue in train drivers in the simulator (Dorrian, Roach, Fletcher, & Dawson, 2007), fatigue in fire-fighters (Smith, Browne, Armstrong, & Ferguson, 2016), alternative technologies for cognitive performance (Brunet, Dagenais, Therrien, Gartenberg, & Forest, 2017; Dorrian *et al*., 2008), and even transcranial direct current stimulation as an alternative fatigue countermeasure (e.g., caffeine) (McIntire, McKinley, Nelson, & Goodyear, 2017).

Honn, Satterfield, McCauley, Caldwell and van Dongen (2016) examined the efforts of workload by comparing a 9-hour duty day with multiple take-offs and landings versus a duty day of equal duration with a single take-off and landing. In their study it was found that objective measures of fatigue using the 10-min PVT and subjective measures of fatigue and sleepiness using the self-reported Samn-Perelli Fatigue Scale (SPS) (see Samn & Perelli, 1982, for review) and Karolinska Sleepiness Scale (KSS) (see Åkerstedt & Gillberg, 1990, for review), respectively, significantly increased fatigue levels. This is further reinforced by both Dorrian, Baulk and Dawson (2011) and Kathner, Wriessnegger, Muller-Putz, Kubler and Halder (2014) who state that high workload levels increase fatigue levels. Conversely, Goel *et al.* (2014) identified that high workload increased subjective self-reported fatigue and sleepiness ratings, regardless of sleep duration. However, it was also found that sleep restriction produced cumulative increases in the number of lapses, as measured by the PVT. However, Volker, Kirchner and Bock (2016) investigated the relationship between subjective and objective measures of fatigue to better understand the complex nature of workplace fatigue and found that fatigue comprises of three independent components; objective physical fatigue, introspective, and extrospective fatigue. The aim of this study was to develop and validate an alternative offline iOS mobile app version of the online 10-min m-PVT i.e., offline 10-min m-PVT to detect levels of fatigue in frontline safety critical workers.

7.3: Methodology

7.3.1: Ethical approval

The study received ethics approval from Cardiff University's Ethics Committee (EC.16.02.09.4464RA3). The study conformed to the seventh amendment of the Declaration of Helsinki 1964 (World Medical Association, 2013) and was in accordance with the Data Protection Act 1998 and the General Data Protection Regulation (GDPR). All train drivers gave their informed written (Appendix OO) as well as electronic consent following the explanation of the nature of the study.

7.3.2: Participants

A total of 40 train drivers (35 male and 5 female) mean age 43 years (SD = 7.4), which represented 6.4 per cent of all drivers across the franchise were voluntary recruited to take part in this study (see Table 6, for demographic characteristics). Drivers were recruited from the two largest depots; Cardiff Valleys (n = 30) and Cardiff Mainline (n = 10), representing 26.5 per cent and 3.4 per cent respectively, of their corresponding depot establishment.

	Depot: Cardiff Valleys (n = 30)			**Depot: Cardiff Mainline (n = 10)**			**Total Across Depots (n = 40)**		
	n	**% Mean**	**SD**	**n**	**% Mean**	**SD**	**n**	**% Mean**	**SD**
Age		41.1	6.5		49.3	6.8		43.2	7.4
Gender									
Male	27	90.0%		8	80.0%		35	87.5%	
Female	3	10.0%		2	20.0%		5	12.5%	
Marital Status									
Single, never married	4	13.3%		1	10.0%		5	12.5%	
Married or domestic partnership	24	80.0%		8	80.0%		32	80.0%	
Separated	1	3.3%		0	0.0%		1	2.5%	
Divorced	1	3.3%		1	10.0%		2	5.0%	
Widowed	0	0.0%		0	0.0%				
Children									
Yes	27	90.0%		8	80.0%		31	77.5%	
No	3	10.0%		2	20.0%		9	22.5%	
Mean Number of Children		1.8	1.6		1.6	1.1		1.8	1.4
Mean Age of Children		11.5	7.9		23.8	7.3		14.2	9.3
Education									
Secondary Education	5	16.7%		4	40.0%		9	22.5%	
Post-Secondary Education	9	30.0%		2	20.0%		11	27.5%	
Vocational Qualification	10	33.3%		4	40.0%		14	35.0%	
Undergraduate Degree	5	16.7%		0	0.0%		5	12.5%	
Post-graduate Degree	1	3.3%		0	0.0%		1	2.5%	
Job Description									
Train Driver	30	100.0%		10	100.0%		40	100.0%	
Driving Instructor	3	10.0%		1	10.0%		4	10.0%	
Ethnicity									
White	30	100.0%		10	100.0%		40	100.0%	
Years certified as a train driver as of 2018		8.4	4.9		15.2	8.7		10.1	6.6
Commuting time to work (minutes)		35	15.1		32	11.3		34	14.2

Table 6: Demographic characteristics

At Arriva Trains Wales (ATW), each depot has a required number of train drivers to be able to operate all operational services at full capacity, which is referred to as the depot establishment (see Table 7). These figures also factor in the additional 27 per cent train drivers required to be able to cover drivers that may be on; annual leave, team brief, sick, no-show, late, incident investigations (i.e., driver being involved in a safety critical incident) as well as training and development. Latest establishment figures as of January 2018, indicate that there was a total of 626 train drivers across all depot establishments. Below are the full number of train drivers at each depot establishment split into both the North and South regions.

Depot Establishment		
	n	**%**
South Wales		
Cardiff Mainline	145	23.2%
Cardiff Valleys	132	21.1%
Carmarthen	59	9.4%
Treherbert	22	3.5%
Rhymney	22	3.5%
North Wales		
Chester	81	12.9%
Crewe	47	7.5%
Machynlleth	34	5.4%
Shrewsbury	30	4.8%
Holyhead	27	4.3%
Llandudno Junction	18	2.9%
Pwllheli	9	1.4%
Total	626	100.0%

Table 7: Depot establishment

It is also important to also point out that there are major differences between train drivers and the undergraduate student population. Firstly, and most noticeably, there were clear age differences between the previous two studies 19.4 (*SD* = 1.6) and 20.0 (*SD* = 2.2), respectively, and train drivers who were on average 43.2 (*SD* = 7.4) years of age, which represented an age difference of ~23 years. However, despite research identifying that reaction time decreases as a result of age (see Woods, Wyma, Yund, Herron, & Reed, 2015, for review), train drivers are selected and recruited on the bases of scoring high on pre-defined vigilance, reaction time, concertation levels, and alertness thresholds. In addition, in the present study, male train drivers represented 87.5 per cent (35 male and 5 female) of the recruited cohort, while in the previous two studies, males only represented

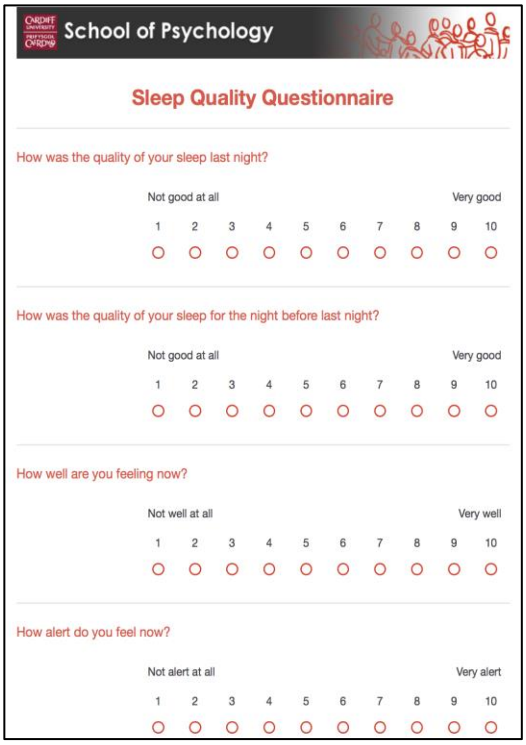
12.2 per cent (9 male and 65 female) and 28.2 per cent (11 male and 28 female), respectively.

7.3.3: Statistical analyses

IBM's Statistical Package for the Social Sciences (SPSS) version 23 for Mac was used to analyse the data. A combination of various statistical procedures were carried out on the data; intraclass correlation coefficient (ICC) analysis, descriptive analyses, Wilcoxon Signed Rank analyses, paired-samples *t*-tests, and analysis of variance (ANOVA). The level of $\alpha < .05$ was used for all statistical tests of this experiment.

7.3.4: Materials and apparatus

Self-reported single-item measures based on the Smith Wellbeing Questionnaire (SWELL) (Smith & Smith, 2017c) were used to compare the impact of sleep quality (see Figure 89) and workload (see Figure 90) for train drivers' high and low workload days. In addition, caffeine consumptions before driving i.e., 'Have you had any caffeine (e.g., coffee, energy drinks, tea, etc.) in the last 24 hours?' and after driving i.e., 'Have you had any caffeine (e.g., coffee, energy drinks, tea, etc.) since starting work today?' for both high and low workload days were also compared. These questions as well as demographic questions were administered on Apple's iPhone 6s Plus device, running on the Qualtrics Surveys (Qualtrics Labs, Inc. version 13.37.02) and powered by JavaScript Form Engine (JFE version 15.03.04).



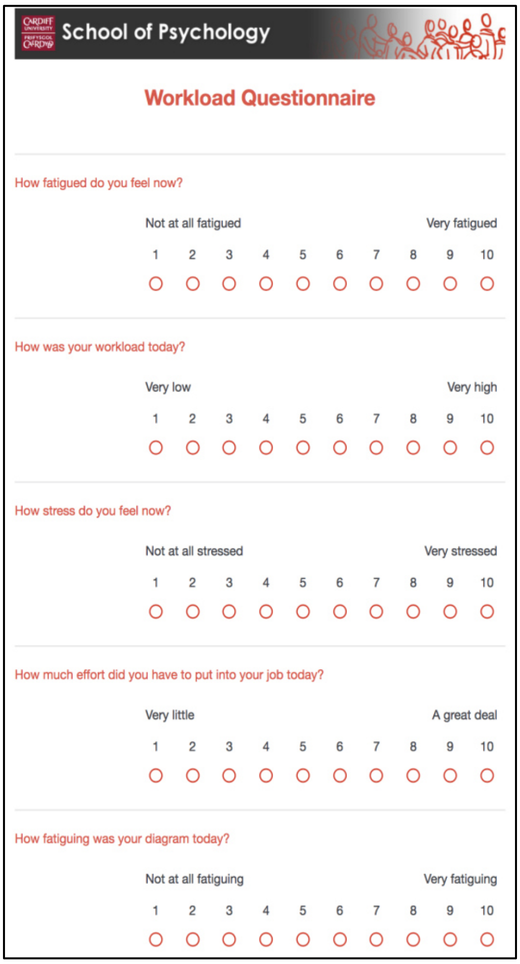


Figure 89: Sleep quality single-item questionnaire

Figure 90: Workload single-item questionnaire

The mobile Psychomotor Vigilance Task (m-PVT) was presented to train drivers on an Apple iPhone 6s Plus mobile device running Apple's iOS version 10.3.2 (Apple Inc.) (see Figure 91). The iPhone 6s Plus had the following hardware configurations; system chip (Apple A9 APL1022), processor (Dual-core, 1840 MHz, Twister, 64-bit), graphics processor (PowerVR GT7600), and system memory (2048 MB RAM). The m-PVT was displayed on either a 7-inch (diagonal) 1280 × 800-pixel (WXGA) native resolution at 216 pixels per inch (ppi) liquid crystal display (LCD) display, that was a 5.5-inch (diagonal) 1920 × 1080-pixel native resolution at 401 ppi Retina high definition display. The m-PVT was developed using Apple's native Objective-C programming language on the graphical interface Xcode environment (version 9.4.1), which is Apple's IDE (Integrated Development Environment) for iOS app development. In order to increase validity and standardisation, all instructions both the high workload day (see Appendix PP, QQ, RR, SS, TT, and UU) and low workload day (see Appendix VV, WW, XX, YY, ZZ, and AAA) were provided to train drivers in written form. Train drivers were also verbally debriefed at the end of the study to explain the nature of the study in accordance with The British Psychological Society (BPS) code of human research ethics (see BPS, 2014, for review) as well as train drivers being provided with a debrief sheet to take home (see Appendix BBB).

Figure 91: iOS App: Mobile Psychomotor Vigilance Task (m-PVT)

7.3.5: Design

The experiment employed a 2 × 2 × 10 three-way repeated analysis of variance (ANOVA) comparing 2 × workload (high workload, or low workload) × 2 time of day (before driving m-PVT, or after driving m-PVT) × 10 time on task (1-minute; 2 minutes; 3 minutes, 4 minutes; 5 minutes; 6 minutes; 7 minutes; 8 minutes; 9 minutes; or 10 minutes).

The study involved train drivers attending four sessions, each lasting approximately 20-minutes; a before driving session (i.e., before booking-on) and an after driving session (i.e., after booking-off) on two different days; one day that train drivers perceived to be a high workload day (i.e., extensive driving shift duration) and another day train drivers perceived to be a low workload day (i.e., short driving shift duration or spare/standby/team brief), in exchange for £20. Drivers were instructed to select any two high and low workload days from their allocated shift roster over the coming weeks/months and were instructed that they did not have to be on consecutive days. This was important as it provided drivers with the flexibility and freedom to identify days from a larger period, since some drivers could go weeks or even months without being assigned a low workload day as expressed by some train drivers prior to running the study. Therefore, due to the nature of this study, there was no order effect as drivers were able to choose a high workload day followed by a low workload day or vice versa based on various factors such as; availability (e.g., prior commitments), feasibility (e.g., catching the last train home), scheduling (e.g., short-term diagram alterations) and practicality (e.g., finishing at Canton depot, which some train drivers found inconvenient, despite the researcher being able to accommodate). The study lasted 120 minutes in total for all four sessions.

7.3.6: Procedure

In order to ensure participants were fully aware of the inclusion and exclusion criteria, all train drivers were approached and verbally explained what the fatigue study involved prior to them taking part. The study was administered using an Apple iPhone 6s Plus mobile device. In order to increase validity and standardisation, all instructions were administered to train drivers in written form. This study consisted of two measurements during each of the four sessions. The first measurement was the mobile Psychomotor Vigilance Task (m-PVT), which was an iOS mobile application version of Dinges and

Powell’s (1985) Psychomotor Vigilance Task. The second measurement were the sleep quality questionnaire pre-driving and the workload questionnaire post-driving that were administered within the Qualtrics Surveys iOS mobile application.

In this mobile Psychomotor Vigilance Task (m-PVT) iOS mobile application version (see Figure 92), participants were presented with on-screen instructions and a button at the end that read ‘Start’. In each trial, participants were shown a black screen background, and at the centre of the screen they would be presented with a large red fixation circle. The red fixation circle (i.e., inter-stimulus interval) would remain on the screen for a randomised duration that lasted between 2 – 10 seconds, which was then followed by a yellow stimulus counter. As soon as the inter-stimulus interval reached the randomised duration, a yellow stimulus counter appeared counting rapidly up in milliseconds from 0 – 5 seconds where it would lapse (i.e., error of omission for 0.5 seconds) and begin the next trial, or until the participant tapped on the screen. Once the participant tapped on the screen, their reaction time (i.e., stimulus) would be displayed for 0.5 seconds. At the end of each trial, an empty black background would appear on-screen for 0.5 seconds. There were 87 trials in total that lasted approximately 10 minutes. Kribbs and Dinges (1994) found that after a maximum of three trials, the practice effect for the PVT was removed. Therefore, the first three trials were classified as practice trials and removed from the final analyses. If participants responded prematurely during any trial (i.e., before the timer commenced counting up), the trial would reset. To also ensure participants were made aware of their premature response, the following message in red was displayed on the centre of the screen, ‘You clicked too early! This trial will be reset.’ A visual illustration of the mobile Psychomotor Vigilance Task (m-PVT) is presented in Figure 93.

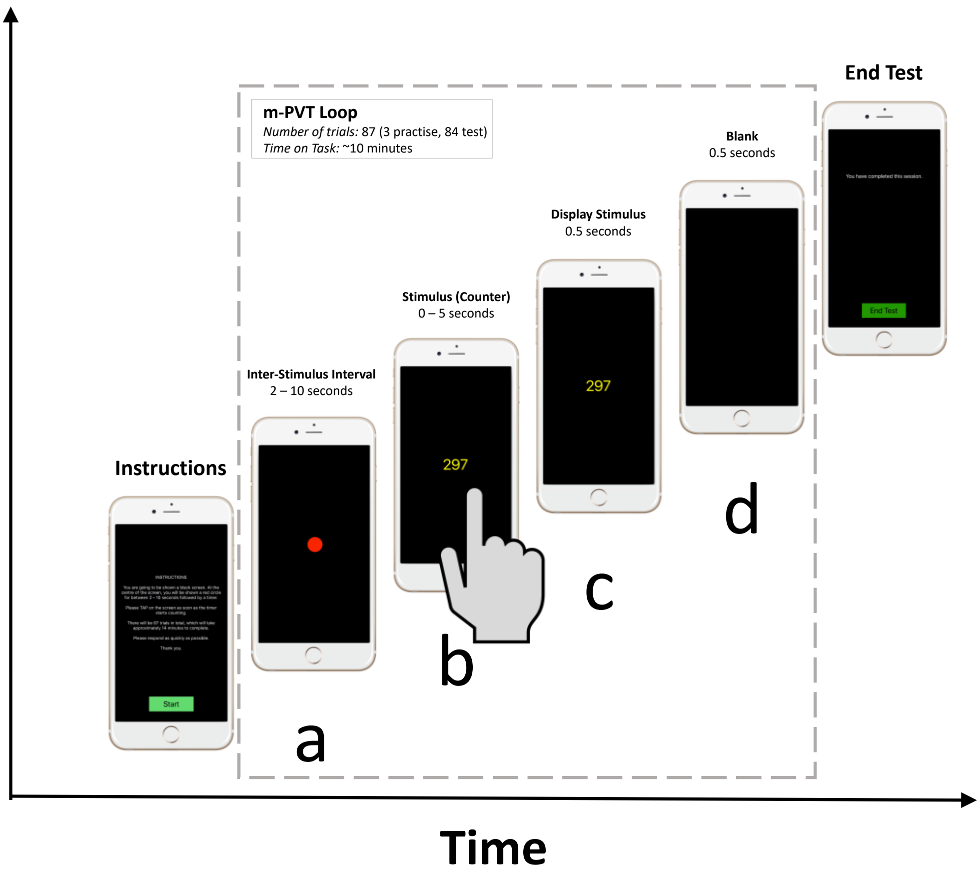


Figure 92: Mobile Psychomotor Vigilance Task (m-PVT) timeline

- **a.** Participants were presented with a large red circle (i.e., inter-stimulus interval), which appeared for a randomised duration between 2 – 10 seconds. If participants responded prematurely, a false start warning message appeared informing them that they clicked too early and that the trial would be reset.
- **b.** As soon as the inter-stimulus interval reached the randomised duration, a yellow stimulus counter appeared counting up in milliseconds from 0 – 5 seconds where it would lapse (i.e., error of omission for 0.5 seconds) and begin the next trial, or until the participant had tapped on the screen.
- **c.** Once the participants had tapped on the screen, their reaction time (i.e., stimulus) would be displayed for 0.5 seconds.
- **d.** At the end of each trial, a black background would appear on-screen for 0.5 seconds.

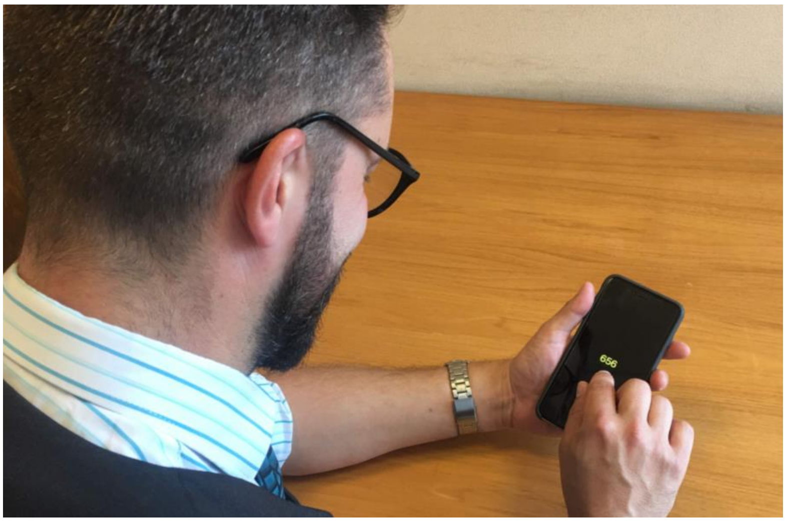


Figure 93: Visual illustration of a train driver using the mobile Psychomotor Vigilance Task (m-PVT)

7.4: Results

The aim of this study was to investigate whether an offline iOS mobile app version of the online 10-min PVT i.e., offline 10-min m-PVT could be used to detect levels of fatigue in frontline safety critical workers. IBM's Statistical Package for the Social Sciences (SPSS) version 23 for Mac was used to analyse the data.

7.4.1: Reliability and validity

The following subsections address the reliability and validly of the offline 10-min m-PVT by highlighting the importance of; reporting the effect size to provide a confidence level in the reported *p*-values, reporting the intraclass correlation coefficient (ICC), which is a widely used reliability analysis. Mauchly's Test of Sphericity, which tests the hypothesis that the variances of the differences between conditions are equal are reported in the following subsections; 7.4.5.1: Mauchly's Test of Sphericity for mean reaction time (RT), 7.4.5.3: Mauchly's Test of Sphericity for mean speed response (1/RT), and 7.4.5.5: Mauchly's Test of Sphericity for mean number of lapses, respectively.

7.4.1.1: Effect size

Chapter 5 outlines the rationale for reporting effect size. Therefore, the following partial eta-squared (η_p^2) threshold criteria were also used in this study; .01 small, .06 medium, and .14 large (Cohen, 1973; Cohen, 1988).

7.4.1.2: Reliability

Chapter 5 outlines the rationale for reporting the intraclass correlation coefficient (ICC). The ICC estimates and their 95% confident intervals based on a mean-rating (k = 20), absolute-agreement, 2-way mixed-effects model revealed that there was an excellent degree of reliability (see Ko & Li, 2016, for review) between all 40 mean reaction times (RTs) i.e., the high workload before driving offline 10-min m-PVT (1-minute, 2 minutes, 3 minutes, 4 minutes, 5 minutes, 6 minutes, 7 minutes, 8 minutes, 9 minutes, and 10 minutes), the high workload after driving offline 10-min m-PVT (1-minute, 2 minutes, 3 minutes, 4 minutes, 5 minutes, 6 minutes, 7 minutes, 8 minutes, 9 minutes, and 10 minutes), the low workload before driving offline 10-min m-PVT (1-minute, 2 minutes, 3 minutes, 4 minutes, 5 minutes, 6 minutes, 7 minutes, 8 minutes, 9 minutes, and 10 minutes), and the low workload after driving offline 10-min m-PVT (1-minute, 2 minutes, 3 minutes, 4 minutes, 5 minutes, 6 minutes, 7 minutes, 8 minutes, 9 minutes, and 10 minutes). The average measure ICC was .990 with a 95% confidence interval from .985 to .994, $F(39, 1521) = 106.10$, $p < .001$.

7.4.2: Diagram duration

A paired-samples t-test was conducted to compare train driver's high workload day diagram duration (i.e., shift duration) with their low workload day diagram duration (see Figure 94). There was a significant difference as presented in minutes between the diagram duration for the high workload day (M = 540.00, SD = 28.86) (i.e., 9 hours 0 minutes) when compared to the low workload day (M = 467.75, SD = 63.66) (i.e., 7 hours

47 minutes) $t(39) = 7.24$, $p < .001$ (2-tailed), $d = 1.48$, indicating a large effect size (Cohen, 1988; Cohen, 1992). This seems to indicate that train drivers were identifying and selecting diagrams that were significantly longer in duration for the high workload day than when compared to diagram durations in the low workload day. However, for the low workload day, these diagrams do not accurately represent train drivers actual diagram duration. This was due in some instances, for example, where train drivers were rostered to work the maximum shift duration of 09:30 but were relieved early of duty at the discretion of the resource management team, which is standard practice when train drivers are rostered the following duties; spare (SP), standby (SB), and team briefing (TB). Therefore, since seven train drivers selected either SP, SB or TB for their low workload day, adjustments were necessary.

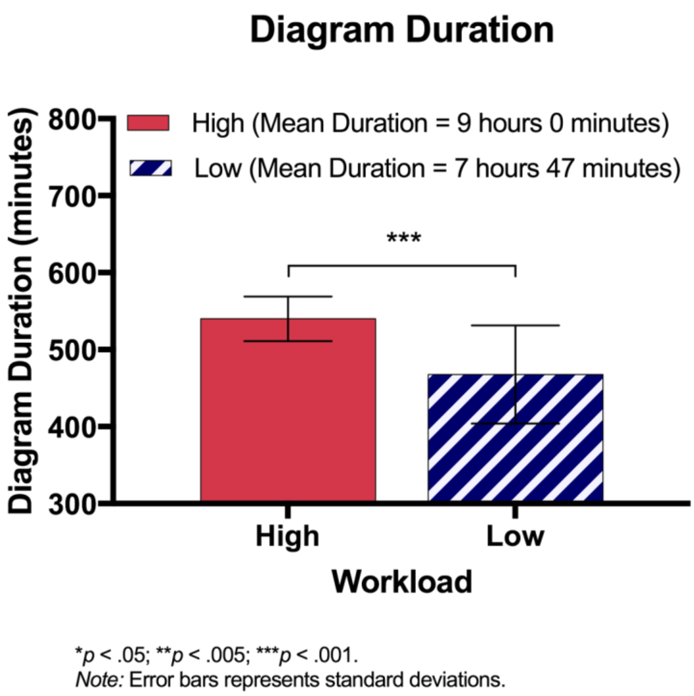


Figure 94: Diagram duration

7.4.2.1: Adjusted diagram duration

A paired-samples *t*-test was conducted to compare train driver's adjusted high workload day diagram duration (i.e., shift duration) with train driver's adjusted low workload day diagram duration (see Figure 95). There was a significant difference as presented in minutes between the diagram duration for the adjusted high workload day (M = 524.85, SD = 28.38) (i.e., 8 hours 44 minutes) when compared to the adjusted low workload day (M = 417.13, SD = 63.62) (i.e., 6 hours 57 minutes) $t(39) = 10.35$, $p < .001$ (2-tailed), d = 2.21, indicating a large effect size (Cohen, 1988; Cohen, 1992). This seems to indicate that train drivers were identifying and selecting diagrams that were significantly longer in duration for the high workload day than when compared to the low workload day even post adjustment for train drivers being relieved early of duty at the discretion of the resource management team.

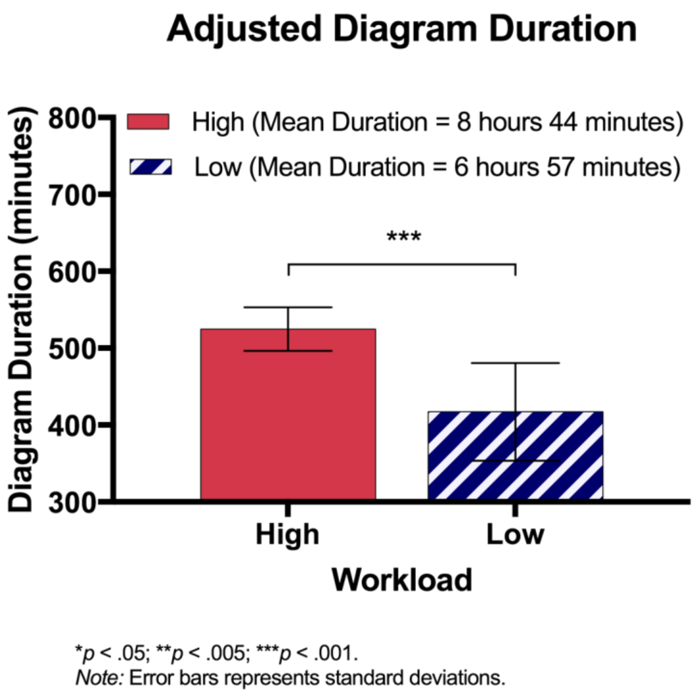


Figure 95: Adjusted diagram duration

7.4.2.2: Comparing diagram duration and adjusted diagram duration for both high and low workload days

Further analyses were conducted to compare train driver's high workload day diagram duration (i.e., shift duration) with train driver's adjusted high workload day diagram duration as well as driver's high workload day diagram duration with train driver's adjusted high workload day diagram duration (see Figure 96). There was a significant difference, as presented in minutes, between the diagram duration for the high workload day (*M* = 540.00, *SD* = 28.86) (9 hours 0 minutes) when compared to the adjusted high workload day (*M* = 524.85, *SD* = 28.38) (i.e., 8 hours 44 minutes) $t(39) = 7.22$, $p < .001$ (2-tailed), $d = 0.54$, indicating a moderate effect size (Cohen, 1988; Cohen, 1992). In addition, there was a significant difference, as presented in minutes, between the diagram duration for the low workload day (*M* = 467.75, *SD* = 63.66) (i.e., 7 hours 47 minutes) when compared to the adjusted low workload day (*M* = 417.13, *SD* = 63.62) (i.e., 6 hours 57 minutes) $t(39) = 3.46$, $p = .001$ (2-tailed), $d = 0.81$, indicating a large effect size (Cohen, 1988; Cohen, 1992).

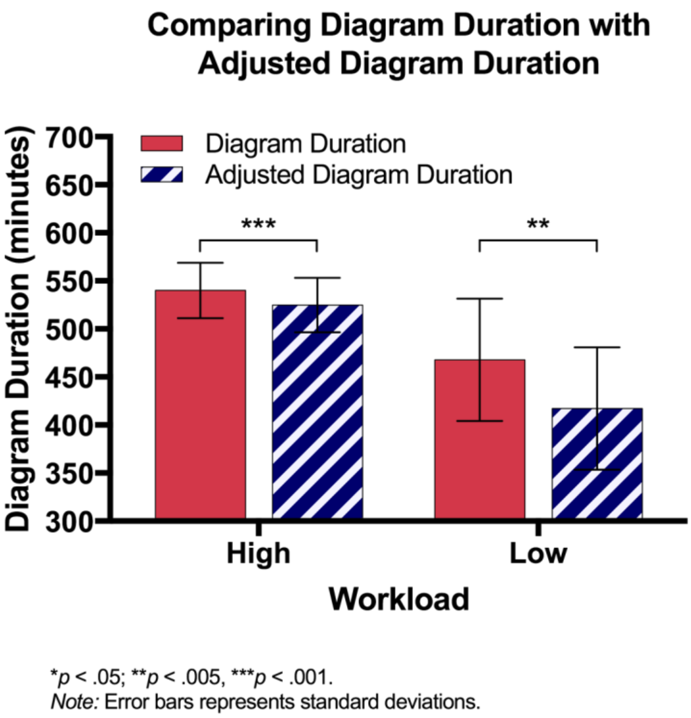


Figure 96: Comparing diagram duration with adjusted diagram duration

These findings seem to indicate that train drivers on average worked 16 minutes less after adjusting their scheduled diagram duration before taking part in the study at the end of their shift for the high workload day. It is important to highlight that train drivers at Arriva Trains Wales (ATW) are provided with an additional 10 minutes after their final diagram turn (i.e., last train journey) to report any issues or safety incidents that may have occurred.

In addition, the 10 minutes is also available for train drivers to return to the driver resource management office, should they need further resource assistance. Therefore, it is common practice for train drivers to finish their diagram (i.e., shift) 10 minutes earlier than their scheduled shift duration, should there be no issues or safety incidents to report. In contrast, on average train drivers worked 50 minutes less after adjusting their scheduled diagram duration before taking part in the study at the end of their shift for the low workload day. This is not too surprising when considering that train drivers identified and selected diagram turns such as; spare (SP), standby (SB), and team briefing (TB) for their low workload day, knowing that there would be a possibility of being relieved early of duty at the discretion of the resource management team, which is standard practice when train drivers are rostered as; SP, SB, and TB. However, it is important to point out that there is a huge discrepancy between the actual scheduled diagram duration for the low workload day when compared to the adjusted diagram duration for the low workload day due to the fact that some drivers were scheduled to work a 9 hours 30 minutes diagram but were relieved of duty significantly earlier.

7.4.3: Consecutive number of days worked

A paired-samples *t*-test was conducted to compare the number of days train drivers worked prior to taking part in the high workload day with the number of days train drivers worked prior to taking part in the low workload day (see Figure 97). There was no significant difference between the number of days train drivers worked prior to taking part in the high workload day (M = 2.85, SD = 1.83) when compared to the number of days train drivers worked prior to taking part in the low workload day (M = 2.75, SD = 1.86) $t(39) = .350$, $p = .728$ (2-tailed), $d = 0.05$, indicating a near zero effect size (Cohen, 1988; Cohen, 1992). This seems to indicate that there was no significant difference in the number of days train drivers worked. In addition, it was found that train drivers on average worked three days prior to taking part in the study, for both the high workload day and low workload day.

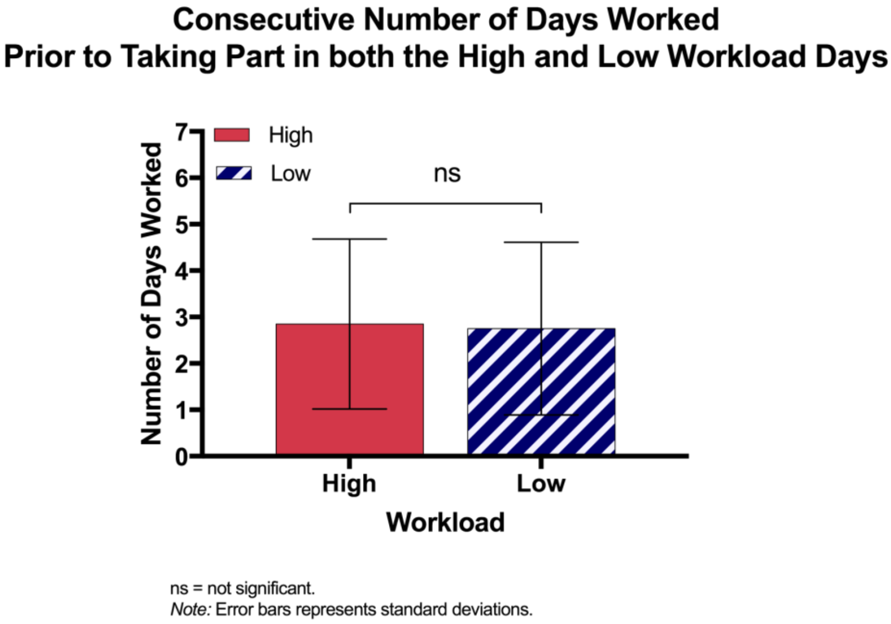
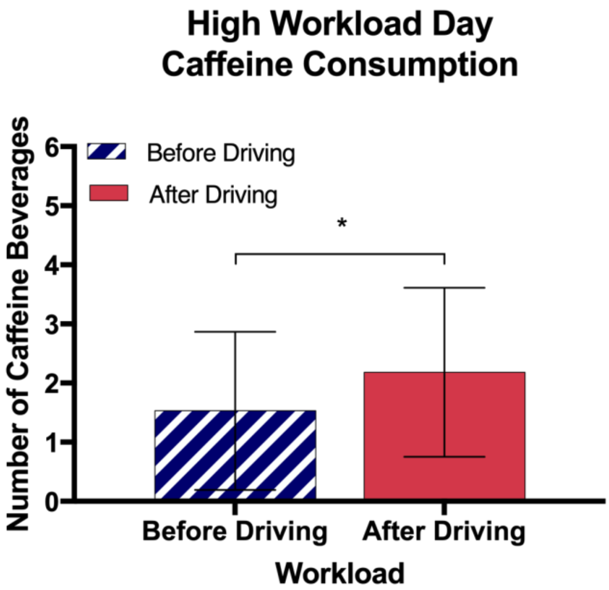


Figure 97: Consecutive number of days worked

7.4.4: Self-reported caffeine consumption

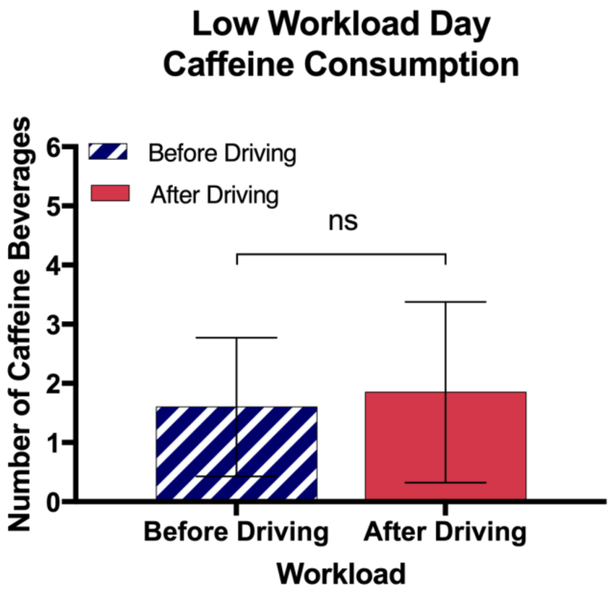
A Wilcoxon Signed Rank Test was conducted to compare train driver's caffeine consumption before driving i.e., 'Have you had any caffeine (e.g., coffee, energy drinks, tea, etc.) in the last 24 hours?' and after driving i.e., 'Have you had any caffeine (e.g., coffee, energy drinks, tea, etc.) since starting work today?' for the high workload day (see Figure 98). The Wilcoxon Signed Rank Test revealed a statistically significant increase in caffeine consumption before driving (*Mdn* = 1) and after driving for the high workload day (*Mdn* = 2), $z = 2.14$, $p = .032$, r = .24, indicating a small effect size (Cohen, 1988).



*p < .05; **p < .005; ***p < .001.
Note: Error bars represents standard deviations.

Figure 98: High workload day caffeine consumption

A Wilcoxon Signed Rank Test was conducted to compare train driver's caffeine consumption before driving and after driving for the low workload day (see Figure 99). The Wilcoxon Signed Rank Test revealed no statistically significant difference in caffeine consumption before driving (*Mdn* = 1) and after driving for the low workload day (*Mdn* = 2), $z = .710$, $p = .478$. These findings identified that train driver's caffeine consumption after driving when compared to before driving was significantly higher for the high workload day, but not for the low workload day, which seems to indicate that perhaps caffeine consumption increases with workload.



ns = not significant.
Note: Error bars represents standard deviations.

Figure 99: Low workload day caffeine consumption

7.4.5: Train Driver’s Fatigue Index (FI) Scores and Risk Index (RI) Scores

The fatigue index (FI) scores and risk index (RI) scores of all 40 train drivers that took part in the study were also extracted from CrewPlan. However, both the FI scores and RI scores for seven drivers were not accurate in CrewPlan for the low workload day. In some instances, this was due to, for example, occasions in which train drivers were rostered to work the maximum shift duration of 9 hours 30 minutes but were relieved early of duty at the discretion of the resource management team, which is standard practice as outlined above when train drivers are rostered the following duties; spare (SP), standby (SB), and team briefing (TB). Therefore, since seven train drivers selected either SP, SB or TB for their low workload day, adjustments were necessary in the form of replacing both their FI scores and RI scores with scores generated by the Health and Safety Executive’s (HSE) Fatigue Risk Index (FRI) calculator for these early releases at the discretion of the resource management team that were not automatically or manually adjusted within the CrewPlan system to reflect train driver’s actual working diagram duration, but instead reflect train driver's allocated shift duration to better align and correspond with the train driver's original roster.

7.4.5.1: Comparing train drivers fatigue index (FI) scores and adjusted fatigue index (AFI) scores for the low workload day

A paired-samples *t*-test was conducted to compare train drivers' fatigue index (FI) scores and adjusted fatigue index (AFI) scores for the low workload day (see Figure 100). There was no significant difference between the FI scores (M = 8.23, SD = 6.29) and the AFI scores for the low workload day (M = 7.74, SD = 5.57) $t(39)$ = 1.04, p = .151 (1-tailed), d = 0.08.

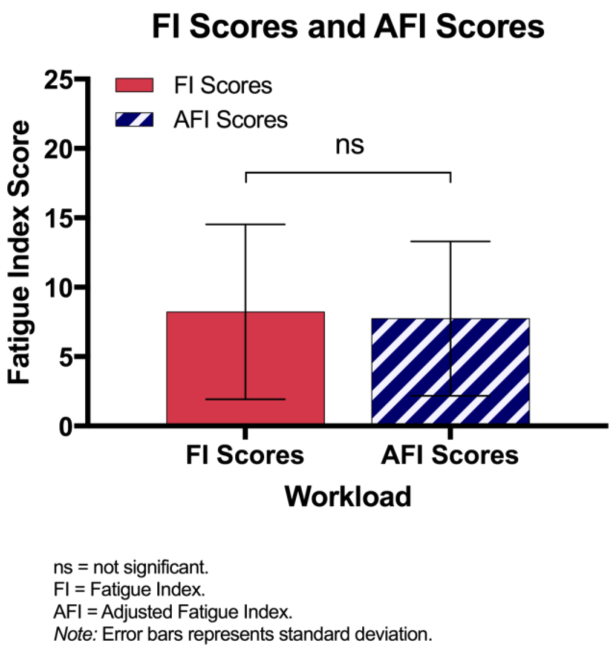
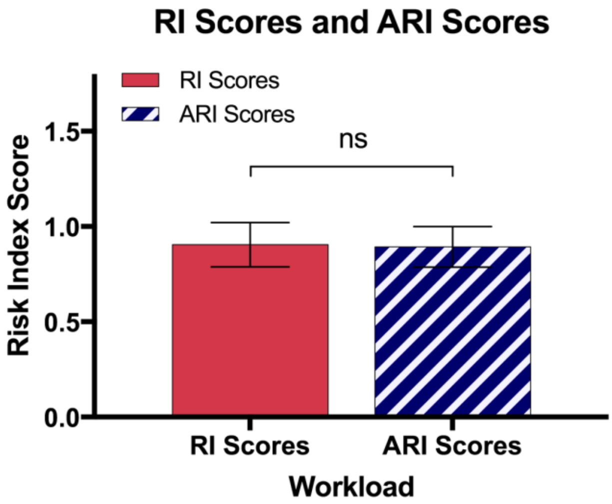


Figure 100: Fatigue index (FI) scores and adjusted fatigue index (AFI) scores for low workload

7.4.5.2: Comparing train drivers risk index (RI) scores and adjusted risk index (ARI) scores for the low workload day

A paired-samples *t*-test was also conducted to compare train drivers' risk index (RI) scores and adjusted risk index (ARI) scores for the low workload day (see Figure 101). There was no significant difference between the RI scores (M = .90, SD = .12) and the ARI scores for the low workload day (M = .89, SD = .11) $t(39)$ = 1.50, p = .071 (1-tailed), d = 0.10, indicating a small effect size (Cohen, 1988; Cohen, 1992). Since there were no train drivers that were relieved of duty at the discretion of the resource management team for the high workload day, no further paired-samples *t*-test was conducted to compare the FI scores and the adjusted FI scores for the high workload day.

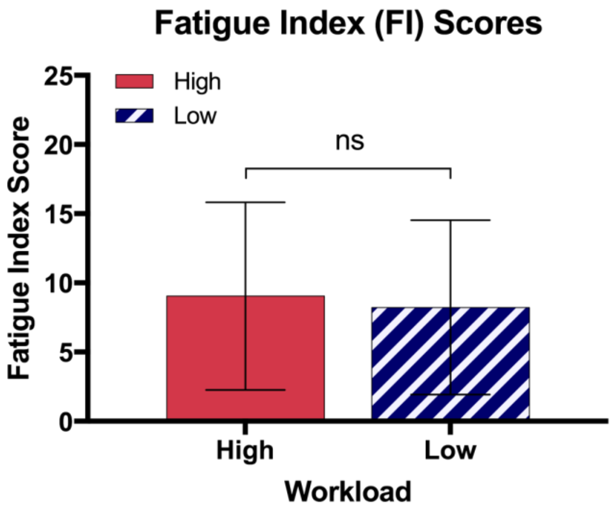


ns = not significant.
FI = Fatigue Index.
AFI = Adjusted Fatigue Index.
Note: Error bars represents standard deviation.

Figure 101: Risk index (RI) scores and adjusted risk Index (ARI) scores for low workload

7.4.5.3: Comparing train drivers fatigue index (FI) scores for both high and low workload days

A paired-samples *t*-test was conducted to compare whether train drivers' fatigue index (FI) scores for the high workload day were significantly higher than the low workload day (see Figure 102). There was no significant difference between the FI scores for the high workload day ($M = 9.04$, $SD = 6.78$) and the low workload day ($M = 8.23$, $SD = 6.29$) $t(39) = .678$, $p = .251$ (1-tailed), $d = 0.12$.

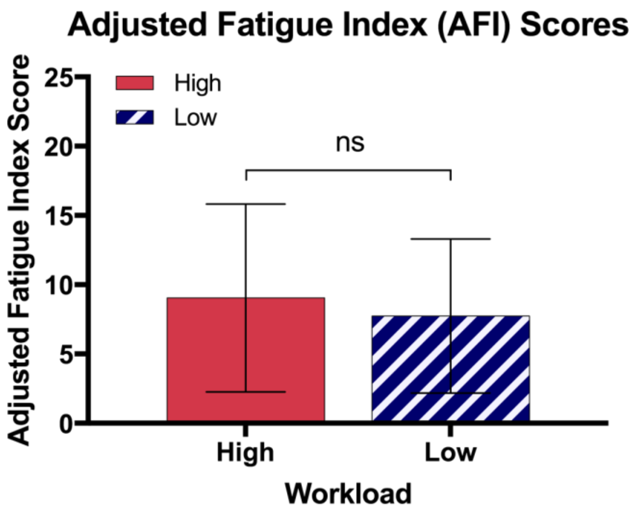


ns = not significant.
Note: Error bars represents standard deviation.

Figure 102: Fatigue index (FI) scores between high workload and low workload

7.4.5.4: Comparing train drivers adjusted fatigue index (AFI) scores for both high and low workload days

A paired-samples *t*-test was conducted to compare whether train drivers' adjusted fatigue index (AFI) scores for the high workload day were significantly higher than the low workload day (see Figure 103). There was no significant difference between the AFI scores for the high workload day (M = 9.04, SD = 6.78) and the low workload day (M = 7.74, SD = 5.57) $t(39)$ = 1.06, p = .149 (1-tailed), d = 0.21.

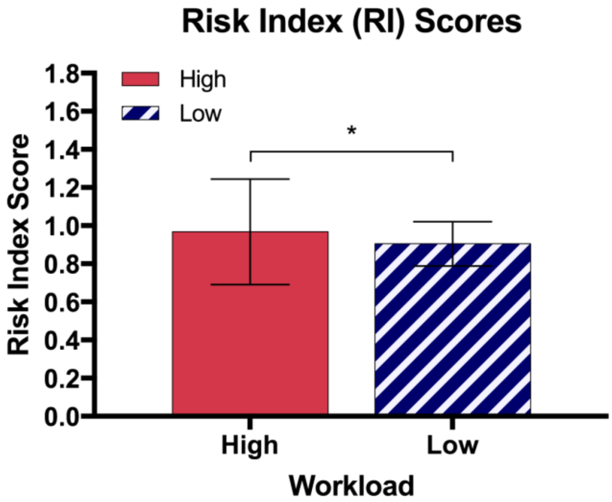


ns = not significant.
Note: Error bars represents standard deviation.

Figure 103: Compering train drivers Adjusted fatigue index (AFI) scores between high workload and low workload

7.4.5.5: Comparing train drivers risk index (RI) scores for both high and low workload days

A paired-samples *t*-test was also conducted to compare whether train drivers' risk index (RI) scores for the high workload day were significantly higher than the low workload day (see Figure 104). There was a significant difference between the RI scores for the high workload day (M = .97, SD = .28) and the low workload day (M = .90, SD = .12) $t(39) = 1.77$, $p = .043$ (1-tailed), $d = 0.28$, indicating a small effect size (Cohen, 1988; Cohen, 1992).

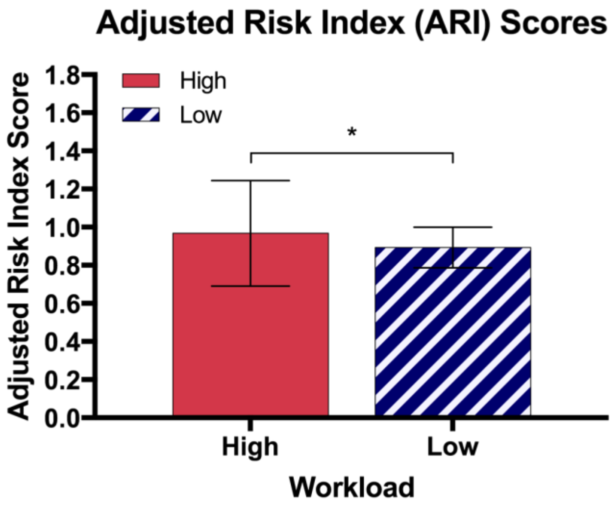


*p < .05.
Note: Error bars represents standard deviation.

Figure 104: Risk index (RI) scores between high workload and low workload

7.4.5.6: Comparing train drivers adjusted risk index (RI) scores for both high and low workload days

A paired-samples *t*-test was also conducted to compare whether train drivers' adjusted risk index (ARI) scores for the high workload day were significantly higher than the low workload day (see Figure 105). There was a significant difference between the ARI scores for the high workload day (M = .97, SD = .28) and the low workload day (M = .98, SD = .11) $t(39) = 2.07$, $p = .002$ (1-tailed), $d = 0.35$, indicating a small effect size (Cohen, 1988; Cohen, 1992).



*$p < .05$.
Note: Error bars represents standard deviation.

Figure 105: Adjusted risk index (ARI) scores between high workload and low workload

7.4.6: mobile Psychomotor Vigilance Task (m-PVT)

A total of 13,920 trials were recorded from 40 train drivers. The first three trials were practice trials and thus were removed from the analyses, leaving 13,400 test trials (84 per session; 336 across all four session). Based on Basner and Dinges (2011) recommendations, test trials with reaction times (RTs) < 100 ms i.e., false start, which accounted for 0.05 per cent (n = 7) of all RTs, and RTs ≥ 500 ms i.e., number of lapses, which accounted for 2.75 per cent (n = 370) of all RTs were excluded from the final analyses, which represented a total of 2.81 per cent (n = 377) of all test trials. All 2.81 per cent (n = 377) of RTs ≥ 500 ms (i.e., number of lapses) were analysed separately.

7.4.6.1: Mauchly's Test of Sphericity for mean reaction time (RT)

The assumption of sphericity was examined using the Mauchly's Test of Sphericity. For the mean reaction time (RT) analyses, all Mauchly's Test of Sphericity indicated that the assumption of sphericity had not been violated; i.e., time on task, $\chi^2(44) = 71.83$, $p = .005$; workload × time on task, $\chi^2(44) = 40.01$, $p = .649$; time of day × time on task, $\chi^2(44) = 53.98$, $p = .149$; and workload × time of day × time on task, $\chi^2(44) = 31.07$, $p = .931$. In addition, Greenhouse–Geisser test was reported for workload, time of day, and workload × time of day, as nothing was known about sphericity.

7.4.6.2: Mean reaction time (RT)

Figure 106 and Figure 107 present the illustrated mean reaction times (RTs) across the different conditions split into the high workload and low workload day. The mean RTs were submitted to a 2 × 2 × 10 three-way repeated analysis of variance (ANOVA) comparing 2 × workload (high workload, or low workload) × time of day (before driving m-PVT, or after driving m-PVT) × 10 time on task (1-minute; 2 minutes; 3 minutes, 4 minutes; 5 minutes; 6 minutes; 7 minutes; 8 minutes; 9 minutes; or 10 minutes).

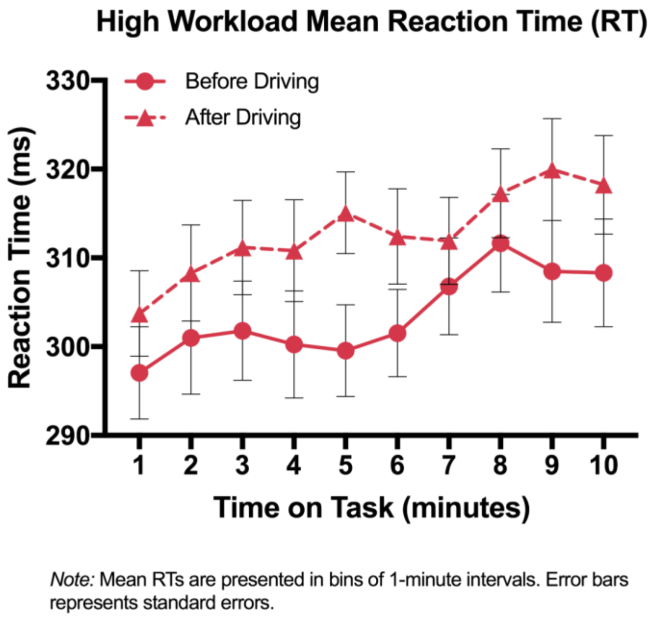


Figure 106: Mean reaction times (RTs) for high workload and time on task (1-minute intervals)

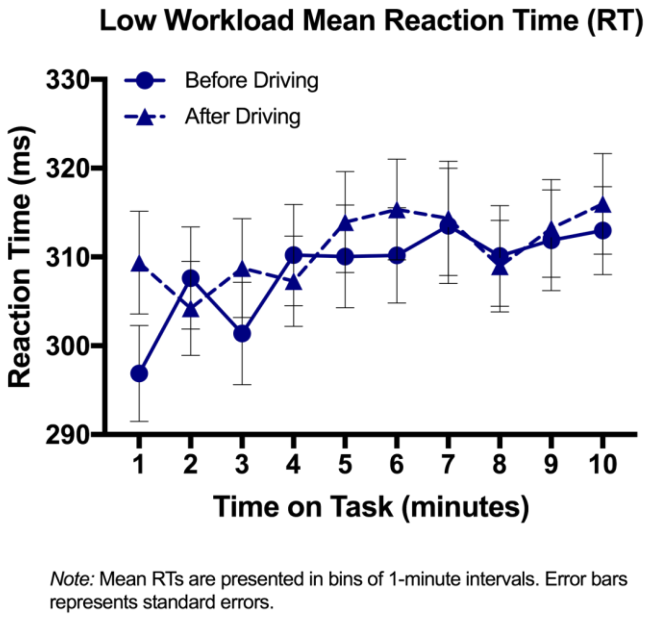


Figure 107: Mean reaction times (RTs) for low workload and time on task (1-minute intervals)

There was no significant main effect of workload, Greenhouse–Geisser = 1.00, $F(1, 39)$, .530, p = .471, η_p^2 = .01. There was a significant main effect of time of day, Greenhouse–Geisser = 1.00, $F(1, 39)$, 10.45, p = .002, η_p^2 = .21, indicating a large effect size (Cohen, 1973; Cohen, 1988; Levine & Hullett, 2002), which provides a high level of confidence in the reported p-value. There was a significant main effect of time on task, Sphericity Assumed, $F(9, 351)$, 9.85, p < .001, η_p^2 = .20, indicating a large effect size (Cohen, 1973; Cohen, 1988; Levine & Hullett, 2002), which also provides a high level of confidence in the reported p-value. There was a significant two-way interaction between workload × time of day, Greenhouse–Geisser = 1.00, $F(1, 39)$, 5.05, p = .030, η_p^2 = .12, indicating a medium effect size (Cohen, 1973; Cohen, 1988; Levine & Hullett, 2002), which provides a moderate level of confidence in the reported p-value. There were no significant two-way interaction between workload × time on task, Sphericity Assumed, $F(9, 351)$, 1.69, p = .090, η_p^2 = .04; as well as a two-way interaction between time of day × time on task, Sphericity Assumed, $F(9, 351)$, 1.43, p = .176, η_p^2 = .04. There was also no significant three-way interaction between workload × time of day × time on task, Sphericity Assumed, $F(9, 351)$, 1.30, p = .236, η_p^2 = .03.

7.4.6.2.1 Mean reaction time (RT) Post-Hoc tests

The main effect of time of day is illustrated in Figure 108 and was followed by post-hoc tests with Bonferroni correction for multiple comparisons. Post-hoc tests showed that train drivers mean RTs were significantly faster in the morning (*M* = 306.06 ms, *SE* = 5.02 ms) when compared to the afternoon (*M* = 312.01 ms, *SE* = 4.62 ms) *p* = .002.

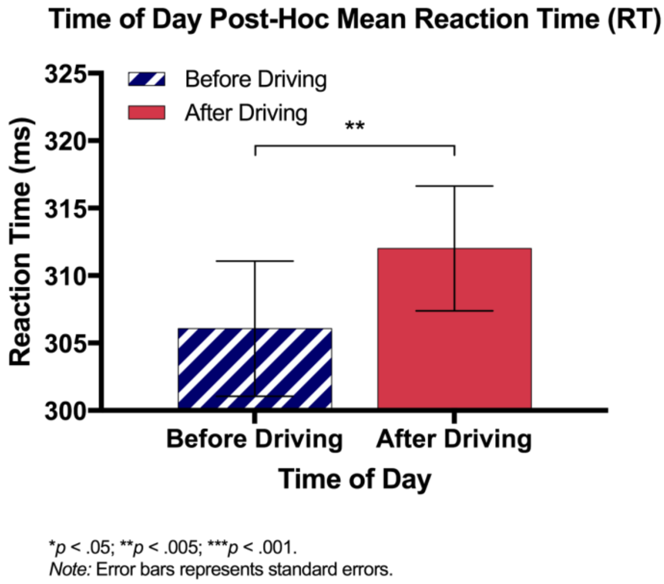


Figure 108: Time of Day

The main effect of time on task is illustrated in Figure 109 and was followed by post-hoc tests with Bonferroni correction for multiple comparisons. Post-hoc tests showed that train drivers' mean RTs were significantly consecutively slower after 5-minutes (*M* = 309.66 ms, *SE* = 4.62 ms) on the m-PVT temporal attention task when compared to mean RTs after the first minute (*M* = 301.76 ms, *SE* = 4.77 ms) *p* = .004. From 5-minutes onwards, all train drivers had consistently slower mean RTs than from the first minute on the m-PVT temporal attention task (*M* = 301.76 ms, *SE* = 4.77 ms) when compared to; 6 minutes (*M* = 309.87 ms, *SE* = 4.75 ms) *p* = .016, 7 minutes (*M* = 311.65 ms, *SE* = 5.02 ms) *p* < .001, 8 minutes (*M* = 312.00 ms, *SE* = 4.77 ms) *p* = .002, 9 minutes (*M* = 313.38 ms, *SE* = 5.02 ms) *p* < .001, and 10 minutes (*M* = 313.88 ms, *SE* = 4.76 ms) *p* < .001.

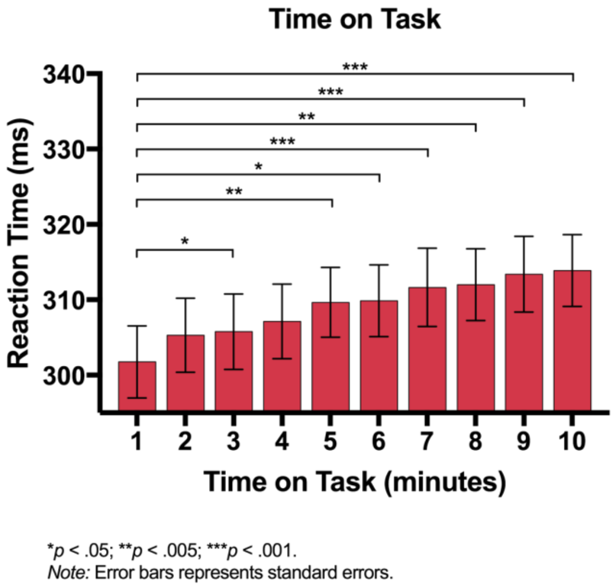
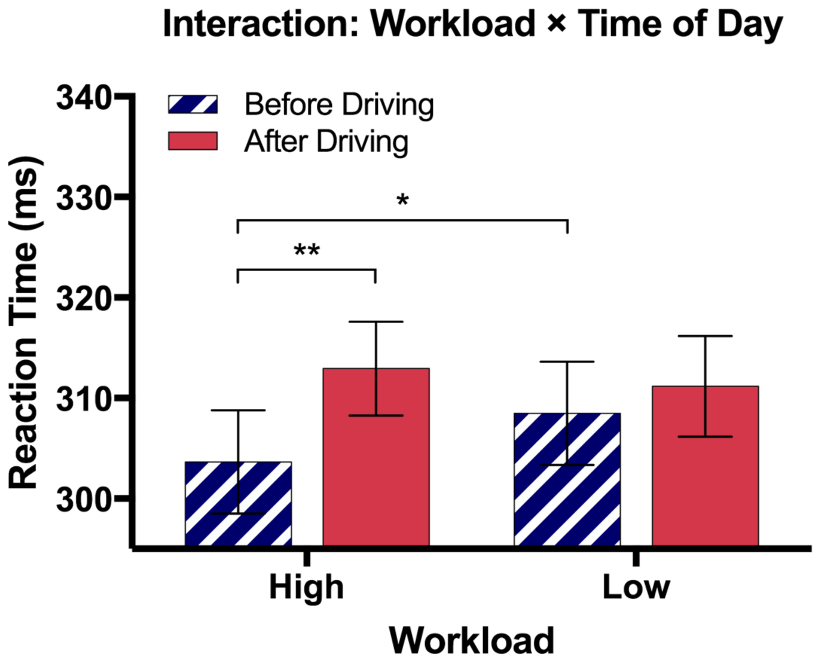


Figure 109: Time on Task

The interaction between workload and time of day is illustrated in Figure 110 and was followed by post-hoc *t*-tests with Bonferroni correction for multiple comparisons. On the high workload day, train drivers had significantly faster mean RTs before driving (M = 303.64 ms, SE = 5.14 ms) when compared to mean RTs after driving (M = 312.89 ms, SE = 4.67 ms), p = .002. This was a statistical decrease of 9.25 ms from train drivers mean RTs before driving and after driving. In comparison, on the low workload day, there was no significant difference in train drivers mean RTs before driving (M = 308.49 ms, SE = 5.14 ms) when compared to mean RTs after driving (M = 311.14 ms, SE = 5.01 ms), p = .159. In contrast, train drivers had significantly faster mean RTs before driving in the high workload day (M = 303.64 ms, SE = 5.14 ms) when compared to mean RTs before driving in the low workload day (M = 308.49 ms, SE = 5.14 ms), p = .036. However, there was no significant difference in train drivers mean RTs after driving in the high workload day (M = 312.89 ms, SE = 4.67 ms) when compared to mean RTs before driving in the low workload day (M = 311.14 ms, SE = 5.01 ms), p = .549.



*p < .05; **p < .005.
Note: Error bars represents standard errors.

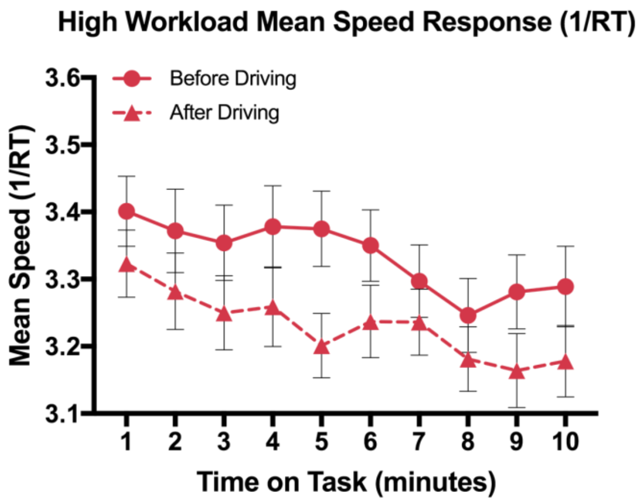
Figure 110: Workload and Time of Day interaction

7.4.6.3: Mauchly's Test of Sphericity for mean speed response (1/RT)

The assumption of sphericity was examined using the Mauchly's Test of Sphericity. For the mean speed response (1/RT) analyses, Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated for time on task, $\chi^2(44) = 67.24$, $p = .014$. Therefore, the Greenhouse-Geisser test was reported instead of Sphericity Assumed, since Greenhouse–Geisser Epsilon was less than .75 for time on task. All other Mauchly's Test of Sphericity indicated that the assumption of sphericity had not been violated: workload × time on task, $\chi^2(44) = 38.79$, $p = .699$; time of day × time on task, $\chi^2(44) = 43.41$, $p = .503$; and workload × time of day × time on task, $\chi^2(44) = 32.92$, $p = .892$. In addition, Greenhouse–Geisser test was reported for workload, time of day, and workload × time of day, as nothing was known about sphericity.

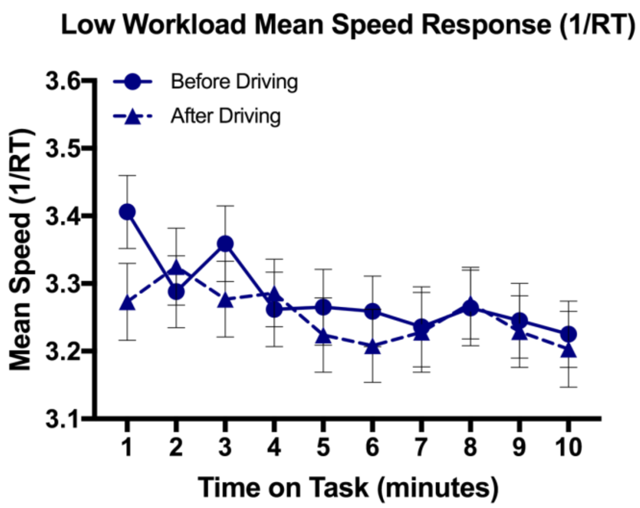
7.4.6.4: Mean speed response (1/RT)

Figure 111 and Figure 112 present the illustrated mean speed responses (1/RTs) across the different conditions split into the high workload and low workload day. The mean 1/RTs were submitted to a 2 × 2 × 10 three-way repeated analysis of variance (ANOVA) comparing 2 × workload (high workload, or low workload) × time of day (before driving m-PVT, or after driving m-PVT) × 10 time on task (1-minute; 2 minutes; 3 minutes, 4 minutes; 5 minutes; 6 minutes; 7 minutes; 8 minutes; 9 minutes; or 10 minutes).



Note: Mean 1/RTs are presented in bins of 1-minute intervals. Error bars represents standard errors.

Figure 111: Mean reaction times (RTs) for high workload and time on task (1-minute intervals)



Note: Mean 1/RTs are presented in bins of 1-minute intervals. Error bars represents standard errors.

Figure 112: Mean reaction times (RTs) for low workload and time on task (1-minute intervals)

There was no significant main effect of workload, Greenhouse-Geisser = 1.00, $F(1, 39)$, .547, $p = .464$, $\eta_p^2 = .01$. There was a significant main effect of time of day, Greenhouse-Geisser = 1.00, $F(1, 39)$, 13.51, $p = .001$, $\eta_p^2 = .26$, indicating a large effect size (Cohen, 1973; Cohen, 1988; Levine & Hullett, 2002), which provides a high level of confidence in the reported *p*-value. There was a significant main effect of time on task, Greenhouse-Geisser = .709, $F(6.38, 248.80)$, 11.14, $p < .001$, $\eta_p^2 = .22$, indicating a large effect size (Cohen, 1973; Cohen, 1988; Levine & Hullett, 2002), which also provides a high level of confidence in the reported *p*-value. There was a significant interaction between workload × time of day, Greenhouse-Geisser = 1.00, $F(1, 39)$, 5.93, $p = .020$, $\eta_p^2 = .13$, indicating a medium effect size (Cohen, 1973; Cohen, 1988; Levine & Hullett, 2002), which provides a moderate level of confidence in the reported *p*-value. There was no significant two-way interaction between workload × time on task, Sphericity Assumed, $F(9, 351)$, 1.71, $p = .086$, $\eta_p^2 = .04$; as well as two-way interaction between time of day × time on task, Sphericity Assumed, $F(9, 351)$, 1.46, $p = .163$, $\eta_p^2 = .04$. There was also no significant three-way interaction between workload × time of day × time on task, Sphericity Assumed, $F(9, 351)$, 1.32, $p = .224$, $\eta_p^2 = .03$.

7.4.6.4.1 Mean speed response (1/RT) Post-Hoc tests

The main effect of time of day is illustrated in Figure 113 and was followed by post-hoc tests with Bonferroni correction for multiple comparisons. Post-hoc tests showed that train drivers mean 1/RTs were significantly quicker in the morning ($M = 3.31$, $SE = .05$) when compared to the afternoon ($M = 3.24$, $SE = .05$) $p = .001$.

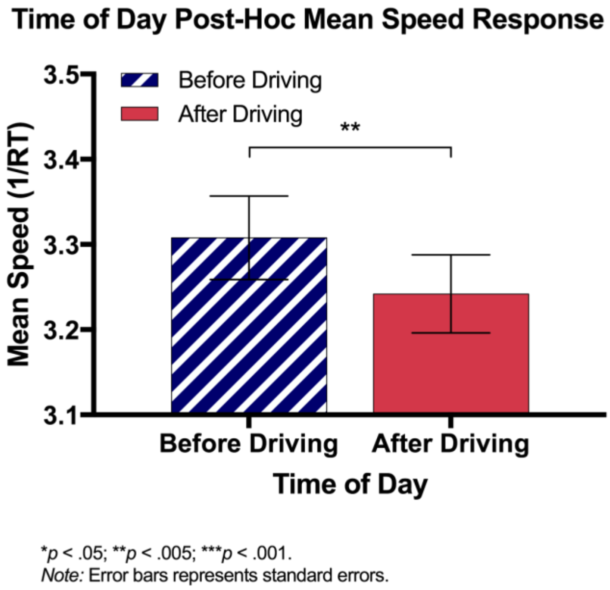
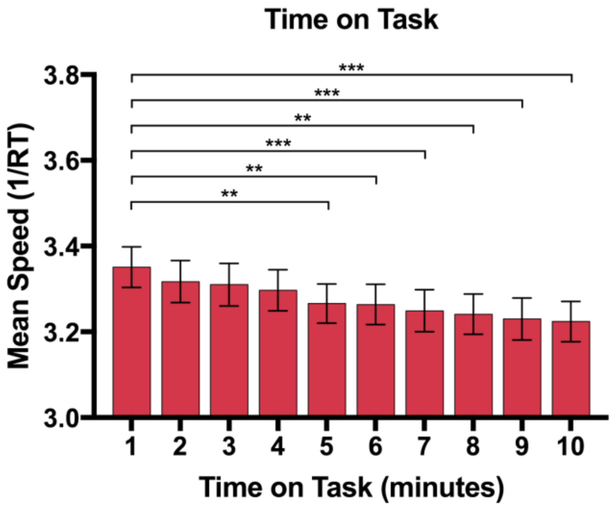


Figure 113: Time of Day

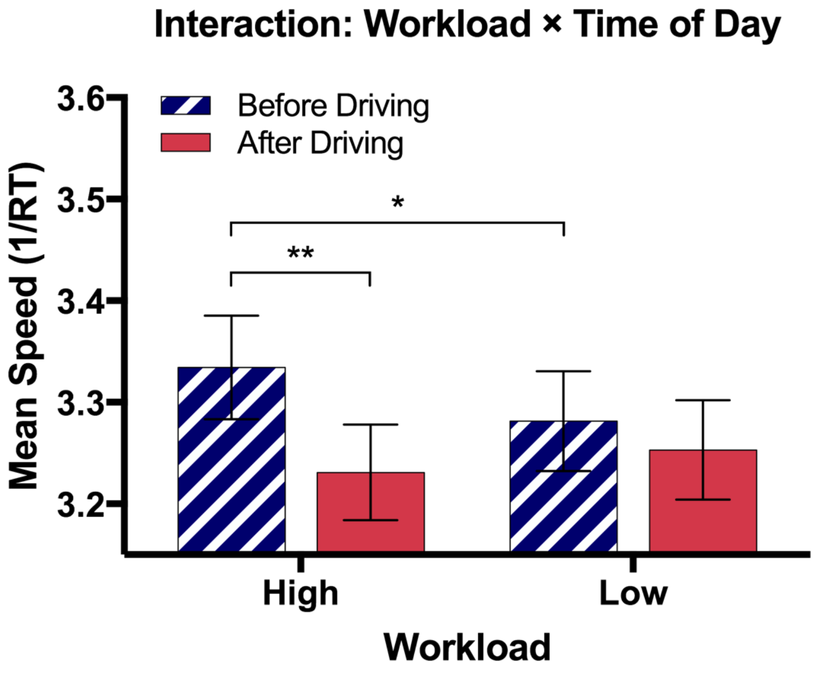
The main effect of time on task is illustrated in Figure 114 and was followed by post-hoc tests with Bonferroni correction for multiple comparisons. Post-hoc tests showed that train drivers mean 1/RTs were significantly consecutively slower after 5-minutes (*M* = 3.27, *SE* =.05) on the m-PVT temporal attention task when compared to mean 1/RTs after the first minute (*M* = 3.35, *SE* = .05) *p* = .003. From 5-minutes onwards, all train drivers had consistently slower mean 1/RTs than from the first minute on the m-PVT temporal attention task (*M* = 3.35, *SE* = .05) when compared to; 6 minutes (*M* = 3.26, *SE* = .05) *p* = .004, 7 minutes (*M* = 3.25, *SE* = .05) *p* < .001, 8 minutes (*M* = 3.24, *SE* = .05) *p* = .001, 9 minutes (*M* = 3.23, *SE* = .05) *p* < .001, and 10 minutes (*M* = 3.22, *SE* = .05) *p* < .001.



*p < .05; **p < .005; ***p < .001.
Note: Error bars represents standard errors.

Figure 114: Time on Task

The interaction between workload and time of day is illustrated in Figure 115 and was followed by post-hoc *t*-tests with Bonferroni correction for multiple comparisons. On the high workload day, train drivers had significantly better mean 1/RTs before driving (M = 3.33, SE = .05) when compared to mean 1/RTs after driving (M =3.23, SE = .05), p = .001. By comparison, on the low workload day, there was no significant difference in train drivers mean 1/RTs before driving (M = 3.28, SE = .05) when compared to mean 1/RTs after driving (M = 3.25, SE = .05), p = .112. In contrast, train drivers had significantly better mean 1/RTs before driving in the high workload day (M = 3.33, SE = .05) when compared to mean 1/RTs before driving in the low workload day (M = 3.28, SE = .05), p = .026. However, there was no significant difference in train drivers mean 1/RTs after driving in the high workload day (M = 3.23, SE = .05 when compared to RTs before driving in the low workload day (M = 3.25, SE = .05), p = .469.



*p < .05; **p < .005, ***p < .001.
Note: Error bars represents standard errors.

Figure 115: Workload and Time of Day interaction

7.4.6.5: Mauchly's Test of Sphericity for mean number of lapses

The assumption of sphericity was examined using the Mauchly's Test of Sphericity. For the mean number of lapses analyses, all Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated: time on task, $\chi^2(44) = 119.45$, $p < .001$; workload × time on task, $\chi^2(44) = 103.80$, $p < .001$; time of day × time on task, $\chi^2(44) = 127.66$, $p < .001$; and workload × time of day × time on task, $\chi^2(44) = 111.02$, $p < .001$. Therefore, the Greenhouse-Geisser test was reported instead of Sphericity Assumed, since Greenhouse–Geisser Epsilon was less than .75 for time on task, workload × time on task, time of day × time on task, and workload × time of day × time on task. In addition, the Greenhouse–Geisser test was also reported for workload, time of day, and workload × time of day, as nothing was known about sphericity.

7.4.6.6: Mean number of lapses

From all test trials, a total of 2.75 per cent (n = 370) RTs ≥ 500 ms (i.e., number of lapses) were submitted for data analyses. Figure 116 and Figure 117 present the illustrated mean number of lapses across the different conditions split into the high workload and low workload day. The mean number of lapses were submitted to a 2 × 2 × 10 three-way repeated analysis of variance (ANOVA) comparing 2 × workload (high workload, or low workload) × time of day (before driving m-PVT, or after driving m-PVT) × 10 time on task (1-minute; 2 minutes; 3 minutes, 4 minutes; 5 minutes; 6 minutes; 7 minutes; 8 minutes; 9 minutes; or 10 minutes).

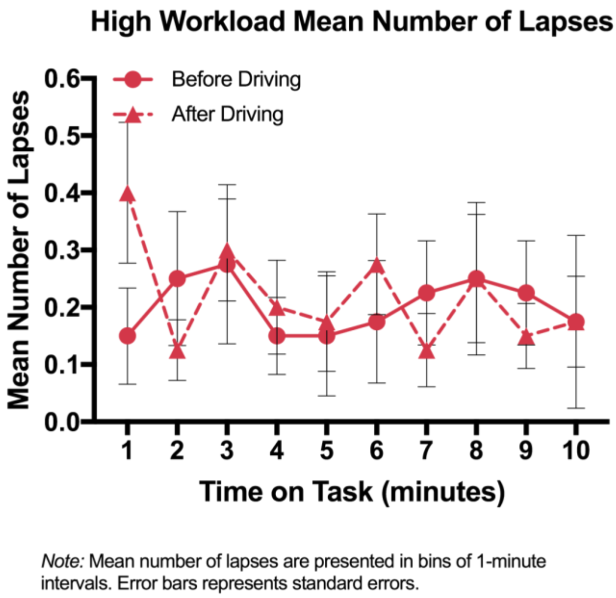
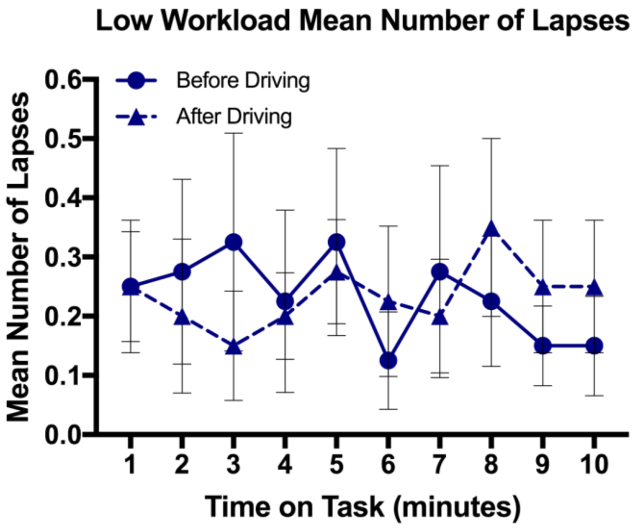


Figure 116: Mean number of lapses for high workload and time on task (1-minute intervals)



Note: Mean number of lapses are presented in bins of 1-minute intervals. Error bars represents standard errors.

Figure 117: Mean number of lapses for low workload and time on task (1-minute intervals)

All main effects and interactions were not significant: workload, Greenhouse-Geisser = 1.00, $F(1, 39)$, .406, $p = .528$, $\eta_p^2 = .01$; time of day, Greenhouse-Geisser = 1.00, $F(1, 39)$, .026, $p = .872$, $\eta_p^2 = .00$; time on task, Greenhouse-Geisser = .521, $F(4.69, 182.73)$, .852, $p = .508$, $\eta_p^2 = .02$; two-way interactions between: workload × time of day, Greenhouse-Geisser = 1.00, $F(1, 39)$, .073, $p = .788$, $\eta_p^2 = .00$; workload × time on task, Greenhouse-Geisser = .613, $F(5.52, 215.32)$, .725, $p = .619$, $\eta_p^2 = .02$; time of day × time on task, Greenhouse-Geisser = .526, $F(4.73, 184.50)$, 1.52, $p = .188$, $\eta_p^2 = .04$; and three-way interactions between workload × time of day × time on task, Greenhouse-Geisser = .558, $F(5.02, 195.89)$, 1.09, $p = .366$, $\eta_p^2 = .03$. Therefore, no further analyses were carried out on the mean number of lapses outcome metrics.

7.4.7: Subjective single-item measures

Wilcoxon Signed Rank tests were carried out to identify whether there were any differences in sleep related single-item measures before driving and workload related single-item measures after driving that could be a significant contributing factor towards any changes in reaction times (RTs) when comparing train driver's high workload day and the low workload day.

7.4.7.1: Before driving single-item measures

As can be seem from Table 8, there was a significant difference in alertness levels before driving for the high workload day (*Md* = 8.0) when compared to the low workload day (*Md* = 7.5) z = -2.55, p = .011, r = .40, indicating a medium effect size of 40 per cent (Cohen, 1988). This seems to indicate that train drivers were more alert before driving on the high workload day than on the low workload day, which could indicate that train drivers are better rested when knowing they would have a more demanding workload ahead (i.e., high workload), which supports the view of train drivers preparing and ensuring they are "fit for duty". There were no other significant single-item measures differences in the: quality of sleep the night before driving for the high workload day (*Md* = 7.0) when compared to the low workload day (*Md* = 7.0) z = -.373, p = .709, r = .06; the quality of sleep the previous night before driving for the high workload day (*Md* = 7.0) when compared to the low workload day (*Md* = 7.0) z = -.098, p = .922, r = .02; and how well train drivers felt before driving for the high workload day (*Md* = 8.0) when compared to the low workload day (*Md* = 8.0) z = -1.42, p = .115, r = .22.

	High Workload		**Low Workload**				
	Median	**SD**	**Median**	**SD**	*z*	***p* (2-tailed)**	***r* (Effect size)**
Before Driving: Sleep Quality Single-Item Measures							
How was the quality of your sleep last night?	7.0	1.8	7.0	1.9	-0.37	= .709	.06
How was the quality of your sleep the night before last night?	7.0	1.8	7.0	1.6	-0.10	= .922	.02
How well are you feeling now?	8.0	1.6	8.0	1.4	-1.42	= .155	.22
How alert do you feel now?	8.0	1.4	7.5	1.6	-2.55	**= .011***	.40

*p < .05; **p < .005; ***p < .001

Table 8: Train driver's self-reported single-item measures of their sleep quality

7.4.7.2: After driving single-item measures

As can be seem from Table 9, all single-item measures after driving between the high workload day compared to the low workload day were statistically significant. Train drivers indicated that they felt more fatigued after driving for the high workload day (*Md*

= 8.0) when compared to the low workload day (*Md* = 6.5) z = -3.08, p = .002, r = .49, indicating a medium effect size of 49 per cent (Cohen, 1988). Train drivers also indicated that their workload was significantly greater after driving for the high workload day (*Md* = 8.0) when compared to the low workload day (*Md* = 5.0) z = -5.26, p < .001, r = .83, indicating a large effect size of 83 per cent (Cohen, 1988). In addition, train drivers felt more stressed after driving for the high workload day (*Md* = 5.0) when compared to the low workload day (*Md* = 3.0) z = -2.97, p = .003, r = .47, indicating a medium effect size of 47 per cent (Cohen, 1988). Furthermore, train drivers felt that they had to put significantly more effort into their job after driving for the high workload day (*Md* = 8.0) when compared to the low workload day (*Md* = 6.0) z = -4.49, p < .001, r = .71, indicating a large effect size of 71 per cent (Cohen, 1988). Moreover, train drivers also reported that their diagram (i.e., shift roster) was more fatiguing for the high workload day (*Md* = 8.0) when compared to the low workload day (*Md* = 4.0) z = -4.91, p < .001, r = .78, indicating a large effect size of 78 per cent (Cohen, 1988). These subjective single-item measures seem to indicate that train drivers selected a significantly more demanding shift for the high workload day than for the low workload day. Moreover, these findings also seem to highlight that the high workload day was; more fatiguing, had a higher workload, was more stressful, required more effort to be put into the job, and had a more fatiguing diagram (i.e., shift roster) than the low workload day.

	High Workload		**Low Workload**				
	Median	**SD**	**Median**	**SD**	*z*	***p* (2-tailed)**	***r* (Effect size)**
After Driving: Workload Single-Item Measures							
How fatigued do you feel now?	8.0	1.7	6.5	2.1	-3.08	**= .002****	.49
How was your workload today?	8.0	1.3	5.0	1.9	-5.26	**< .001*****	.83
How stressed do you feel now?	5.0	2.0	3.0	1.9	-2.97	**= .003****	.47
How much effort did you have to put into your job today?	8.0	1.5	6.0	1.9	-4.49	**< .001*****	.71
How fatiguing was your diagram today?	8.0	1.7	4.0	2.1	-4.91	**< .001*****	.78

Table 9: Train driver's self-reported single-item measures of their workload

7.4.8: Relationships between train driver's subjective and objective measures

Spearman's rank order correlations were run to determine the relationship between train drivers' subjective single-item measures i.e., sleep quality questionnaire and workload questionnaire and objective measures i.e., mobile Psychomotor Vigilance Task (m-PVT), diagram duration as well as the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator as part of Arriva Trains Wales (ATW) Fatigue Risk Management System (FRMS) for both the high workload day (see Table 10) and low workload day (see Table 11)

7.4.8.1: Bivariate analyses: High workload day

All subjective single-item measures of the sleep quality questionnaire (i.e., how was the quality of your sleep last night?, how was the quality of your sleep the night before last night?, how well are you feeling now? and how alert do you feel now?) showed significant positive correlations with each other for the low workload day (see Table 10).

Subjective single-item measures of the workload questionnaire (i.e., how fatigued do you feel now?, how was your workload today?, how stressed do you feel now?, how much effort did you have to put into your job today?, and how fatiguing was your diagram today?), which were administered after driving also showed significant positive correlations with each other, with the exception of two relationships; the association between perceived workload levels and perceived levels of stress, as well as the association between how stressed train drivers felt after driving and how much effort train drivers felt had to be put into their job.

High Workload Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Subjective Single-Item Measures																							
Before Driving (Sleep Quality Questionnaire)																							
1. How was the quality of your sleep last night?	.																						
2. How was the quality of your sleep the night before last night?	.331*	.																					
3. How well are you feeling now?	.527***	.587***	.																				
4. How alert do you feel now?	.573***	.457**	.768***	.																			
Before Driving Caffeine Consumption																							
5. Have you had any caffeine (e.g., coffee, energy drinks, tea, etc.) in the last 24 hours?	-.048	.050	.169	-.104	.																		
After Driving (Workload Questionnaire)																							
6. How fatigued do you feel now?	-.334*	.029	-.172	-.156	-.140	.																	
7. How was your workload today?	-.047	.324*	.185	.212	-.147	.505**	.																
8. How stressed do you feel now?	-.301	-.121	-.381*	-.320*	-.047	.484**	.086	.															
9. How much effort did you have to put into your job today?	.050	.345*	.248	.253	-.034	.484**	.598***	.173	.														
10. How fatiguing was your diagram today?	-.176	.027	-.009	-.014	.008	.646***	.666***	.370*	.476**	.													
After Driving Caffeine Consumption																							
11. Have you had any caffeine (e.g., coffee, energy drinks, tea, etc.) since starting work today?	.007	.216	.278	.102	.092	.006	.348*	-.016	.210	.127	.												
Objective Measures																							
Before Driving																							
mobile-Psychomotor Vigilance Task (m-PVT)																							
12. m-PVT Mean Reaction Time (RT) difference in RT between 1-min and 10-min	-.112	.106	.044	.067	.188	.216	.256	.172	.432**	.186	-.041	.											
13. m-PVT Mean Speed Response (1/RT) difference in 1/RT between 1-min and 10-min	.104	-.098	-.017	-.038	-.152	-.186	-.237	-.155	-.436**	-.163	.030	-.990***	.										
14. m-PVT Mean Number of Lapses (Lapses) difference in Lapses between 1-min and 10-min	-.130	.038	-.041	-.161	-.066	-.181	-.207	-.127	-.095	-.102	.174	-.188	.142	.									
After Driving																							
mobile-Psychomotor Vigilance Task (m-PVT)																							
15. m-PVT Mean Reaction Time (RT) difference in RT between 1-min and 10-min	-.076	-.173	-.213	-.182	.185	-.181	-.312	.113	-.338*	-.110	-.249	-.115	.141	.005	.								
16. m-PVT Mean Speed Response (1/RT) difference in 1/RT between 1-min and 10-min	.076	.182	.202	.177	-.217	.195	.329*	-.120	.333*	.117	.239	.120	-.141	-.014	-.992***	.							
17. m-PVT Mean Number of Lapses (Lapses) difference in Lapses between 1-min and 10-min	-.025	-.116	.223	.134	.255	-.043	-.160	-.218	.153	-.046	.117	-.045	.048	.402*	-.127	.098	.						
Diagram Turn																							
18. Diagram Duration (DD)	-.039	.101	.018	-.005	-.127	-.075	.389*	.033	.045	.195	.093	.073	-.070	-.160	-.190	.201	-.235	.					
19. Adjusted Diagram Duration (ADD)	.045	.031	-.022	.071	-.180	-.104	.184	.040	.018	.054	-.024	.004	-.011	-.124	-.312	.316*	-.307	.766***	.				
Fatigue Risk Index (FRI) Scores																							
20. Fatigue Index (FI)	-.075	-.044	-.169	-.083	-.227	.280	.172	.102	.093	.445**	.054	-.165	.188	.155	.222	-.168	.079	-.161	-.118	.			
21. Risk Index (RI)	-.069	.034	-.045	.034	-.016	.323*	.368*	.131	.286	.519**	.249	.089	-.063	.141	.067	-.013	.250	.117	-.075	.648***	.		
22. Adjusted Fatigue Index (AFI)	-.075	-.044	-.169	-.083	-.227	.280	.172	.102	.093	.445**	.054	-.165	.188	.155	.222	-.168	.079	-.161	-.118	1.000***	.648***	.	
23. Adjusted Risk Index (ARI)	-.069	.034	-.045	.034	-.016	.323*	.368*	.131	.286	.519**	.249	.089	-.063	.141	.067	-.013	.250	.117	-.075	.648***	1.000***	.648**	.

*p < .05; **p < .005; ***p < .001.

Table 10: Spearman’s rank relationship for the high workload day

Low Workload Day	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**
Subjective Single-Item Measures																							
Before Driving (Sleep Quality Questionnaire)																							
1. How was the quality of your sleep last night?	.																						
2. How was the quality of your sleep the night before last night?	**.337***	.																					
3. How well are you feeling now?	**.469****	**.359***	.																				
4. How alert do you feel now?	**.374***	**.316***	**.692****	.																			
Before Driving Caffeine Consumption																							
5. Have you had any caffeine (e.g., coffee, energy drinks, tea, etc.) in the last 24 hours?	-.015	.107	.117	.116	.																		
After Driving (Workload Questionnaire)																							
6. How fatigued do you feel now?	**-.334***	-.249	**-.349***	**-.397***	.076	.																	
7. How was your workload today?	-.063	-.291	-.191	-.014	.062	**.312***	.																
8. How stressed do you feel now?	-.152	-.144	**-.471****	-.068	.004	**.342***	.265	.															
9. How much effort did you have to put into your job today?	.273	-.196	-.072	.046	.011	.113	**.557*****	**.379***	.														
10. How fatiguing was your diagram today?	.020	-.158	-.061	-.028	.115	**.340***	**.805*****	.266	**.558*****	.													
After Driving Caffeine Consumption																							
11. Have you had any caffeine (e.g., coffee, energy drinks, tea, etc.) since starting work today?	.017	-.157	-.148	-.059	.137	-.081	-.175	.033	-.006	**-.361***	.												
Objective Measures																							
Before Driving																							
mobile-Psychomotor Vigilance Task (m-PVT)																							
12. m-PVT Mean Reaction Time (RT) difference in RT between 1-min and 10-min	-.093	.127	.021	.079	.128	.157	-.051	-.062	.069	-.074	-.032	.											
13. m-PVT Mean Speed Response (1/RT) difference in 1/RT between 1-min and 10-min	.017	-.110	-.024	-.070	-.101	-.130	.098	.075	-.068	.119	-.008	**-.984*****	.										
14. m-PVT Mean Number of Lapses (Lapses) difference in Lapses between 1-min and 10-min	.054	.178	.210	.269	.072	-.077	-.074	-.187	.011	-.015	-.205	**.399***	**-.368***	.									
After Driving																							
mobile-Psychomotor Vigilance Task (m-PVT)																							
15. m-PVT Mean Reaction Time (RT) difference in RT between 1-min and 10-min	-.063	.043	.279	.133	.024	.194	-.229	-.216	-.177	-.019	-.141	.070	**-.074**	.173	.								
16. m-PVT Mean Speed Response (1/RT) difference in 1/RT between 1-min and 10-min	.028	-.060	-.275	-.133	.021	-.175	.238	.191	.173	.045	.107	-.051	.061	-.190	**-.988*****	.							
17. m-PVT Mean Number of Lapses (Lapses) difference in Lapses between 1-min and 10-min	-.112	-.019	-.078	-.122	.033	-.004	-.212	-.108	**-.358***	-.189	.200	-.176	.211	.010	-.022	.020	.						
Diagram Turn																							
18. Diagram Duration (DD)	.092	-.015	-.071	.030	**-.362***	.187	-.076	.247	.050	-.242	.062	.194	-.225	-.075	.000	-.045	.134	.					
19. Adjusted Diagram Duration (ADD)	.046	.052	-.199	-.071	-.103	-.126	.137	-.109	-.015	.057	.124	-.037	.052	.145	-.136	.101	.141	.039	.				
Fatigue Risk Index (FRI) Scores																							
20. Fatigue Index (FI)	.093	.095	-.214	-.132	**-.447***	.020	-.029	.005	.009	-.202	.044	.098	-.121	.144	.029	-.049	.110	**.398***	.308	.			
21. Risk Index (RI)	-.054	.120	-.205	-.167	-.050	-.076	**-.347***	-.069	-.226	**-.429***	-.082	.280	-.300	-.031	-.042	.063	.106	.311	.057	.280	.		
22. Adjusted Fatigue Index (AFI)	-.112	.146	**-.352***	-.124	-.080	.131	-.104	.038	-.211	-.118	.024	.120	-.157	.118	.107	-.121	-.018	.166	**.351***	.178	**.334***	.	
23. Adjusted Risk Index (ARI)	-.135	-.170	-.305	-.015	.098	.038	-.035	.072	.126	-.053	.077	.072	-.095	.094	.058	-.057	.073	.134	.290	-.038	**.484****	**.648*****	.

*p < .05; **p < .005; ***p < .001.

Table 11: Spearman's rank relationship for the low workload day

There were no significant correlations between caffeine consumption before driving i.e., 'Have you had any caffeine (e.g., coffee, energy drinks, tea, etc.) in the last 24 hours?' and any other subjective as well as objective measures. However, there was a significant positive correlation between caffeine consumption after driving i.e., 'Have you had any caffeine (e.g., coffee, energy drinks, tea, etc.) since starting work today?' and perceived workload levels $r(40) = .348$, $p = .028$, with an $R^2 = .121$. All other correlations were not significant between caffeine consumption after driving and all other subjective as well as objective measures.

There was a significant negative correlation between train driver's quality of sleep the night before their shift started and perceived fatigue levels after driving $r(40) = -.334$, $p = .035$, with an $R^2 = .112$. All other correlations were not significant between train driver's quality of sleep the night before their shift started and perceived levels of; workload, stress, how much effort train drivers had to put into their job, and how fatiguing the diagram was for train drivers. However, there was a significant positive correlation between the quality of sleep the night prior before their shift started and perceived workload levels $r(40) = .324$, $p = .042$, with an $R^2 = .105$, as well as a significant positive correlation between the quality of sleep the night prior before their shift started and how much effort train drivers had to put into their job $r(40) = .345$, $p = .029$, with an $R^2 = .119$. All other correlations were not significant between train driver's quality of sleep the night prior to the night before their shift started and perceived levels of; fatigue, stress and how fatiguing the diagram was for train drivers.

There was also a significant negative correlation between how train drivers were feeling before their shift started and how stressed train drivers felt after driving $r(40) = -.381$, $p = .015$, with an $R^2 = .145$. All other correlations were not significant between how train drivers were feeling before their shift started and perceived levels of; fatigue, workload, how much effort train drivers had to put into their job, and how fatiguing the diagram was for train drivers.

There was also a significant negative correlation between how alert train drivers felt before their shift started and how stressed train drivers felt after driving $r(40) = -.320$, $p = .044$, with an $R^2 = .102$. All other correlations were not significant between how alert train drivers felt before their shift started and levels of; fatigue, workload, how much effort train drivers had to put into their job, and how fatiguing the diagram was for train drivers.

There were no significant correlations between all the subjective single-item measures of sleep quality and all the objective measures both before driving and after driving using the offline mobile Psychomotor Vigilance Task (m-PVT) i.e., mean reaction times (RTs), mean speed responses (1/RTs), and mean number of lapses, as well as from Arriva Trains Wales' (ATW) fatigue risk management system (FRMS) – CrewPlan i.e., diagram duration (DD), adjusted diagram duration (ADD), including the fatigue index (FI), risk index (RI), adjusted fatigue index (AFI), and adjusted risk index (ARI) scores.

There was a significant positive correlation between train drivers' fatigue levels after driving and ATW's risk index (RI) scores $r(40) = .323$, $p = .042$, with an $R^2 = .104$, as well as a significant positive correlation between train drivers' fatigue levels after driving and the adjusted risk index (RI) scores $r(40) = .323$, $p = .042$, with an $R^2 = .104$. All other correlations were not significant between train drivers' fatigue levels after driving and the other objective measures after driving; m-PVT RTs, 1/RTs, lapses, ATW's DD, ADD, FI, and AFI scores. These findings are consistent with those found in Chapter 3, which identified that based on ATW's Fatigue Risk Management System (FRMS) recommended thresholds of FI score = 45 and RI score = 1.6 (ATW, 2017a), it was found that only 0.2 per cent (n = 1) of train drivers had exceeded the FI score of 45, while 2.1

per cent (n = 12) of train drivers had exceeded the RI score of 1.6, which seems to further support that perhaps the RI score may provide a better predictor in identifying train driver's level of fatigue than the FI score of the HSE's Fatigue Risk Index (FRI) calculator.

There was a significant positive correlation between workload levels and train drivers' m-PVT mean speed responses (1/RTs) $r(40) = .329$, $p = .038$, with an $R^2 = .108$. In addition, there was also a significant positive correlation between workload levels and diagram duration (DD). Moreover, there was a positive correlation between workload levels and ATW's risk index (RI) scores $r(40) = .368$, $p = .020$, with an $R^2 = .135$, as well as a significant positive correlation between workload levels and the adjusted risk index (RI) scores $r(40) = .368$, $p = .020$, with an $R^2 = .135$. All other correlations were not significant between workload levels after driving and the other objective measures after driving; m-PVT RTs, lapses, ADD, FI, AFI, scores.

There were no significant correlations between how stressed train drivers felt after driving and all the other objective measures after driving; m-PVT RTs, 1/RTs, lapses, ATW's DD, ADD, FI, RI, AFI, and ARI scores.

There was a significant negative correlation between how much effort train drivers had to put into their job and train drivers' m-PVT mean reaction times (RTs) $r(40) = -.338$, $p = .033$, with an $R^2 = .104$. There was also a positive correlation between how much effort train drivers had to put into their job and train driver's m-PVT speed responses (1/RTs) $r(40) = -.333$, $p = .036$, with an $R^2 = .111$. All other correlations were not significant between how much effort train drivers had to put into their job and the other objective measures after driving; m-PVT lapses, ATW's DD, ADD, FI, RI, AFI, and ARI scores.

There were significant positive correlations between how fatiguing the diagram was for train drivers and all HSE's Fatigue Risk Index (FRI) calculator scores; the fatigue index (FI) $r(40) = .445$, $p = .004$, with an $R^2 = .198$; the risk index (RI) $r(40) = .519$, $p = .001$, with an R^2 = .269; the adjusted fatigue index (AFI) $r(40) = .445$, $p = .004$, with an R^2 = . 198; and the adjusted risk index (ARI) $r(40) = .519$, $p = .001$, with an R^2 = .269. All other correlations were not significant between how fatiguing the diagram was for train drivers and the other objective measures after driving; m-PVT RTs, 1/RTs, lapses, ATW's DD, and ADD.

7.4.8.2: Bivariate analyses: Low workload day

All subjective single-item measures of the sleep quality questionnaire showed significant positive correlations with each other for the low workload day, as also seen in the high workload day (see Table 11).

Subjective single-item measures of the workload questionnaire also showed significant positive correlations with each other, with the exception of three relationships; the association between fatigue levels and how much effort train drivers felt had to be put into their job, the association between workload levels and how stressed train drivers felt after driving, as well as the association between how stressed train drivers felt after driving and how fatiguing the diagram was for train drivers.

There was a significant negative correlation between caffeine consumption before driving and diagram duration (DD) $r(40) = -.362$, $p = .022$, with an R^2 = .131. In addition, there was also a significant negative correlation between caffeine consumption before driving and ATW's fatigue index (FI) scores (FI) $r(40) = -.447$, $p = .004$, with an R^2 = .20. All other correlations were not significant between caffeine consumption before driving and all other subjective as well as objective measures. Moreover, there was a significant negative correlation between caffeine consumption after driving and how fatiguing the diagram was for train drivers $r(40) = -.361$, $p = .022$, with an R^2 = .13. All other correlations were not significant between caffeine consumption after driving and all other subjective as well as objective measures.

There was a negative correlation between how train drivers were feeling before their shift started and the adjusted fatigue index (AFI) scores after driving $r(40) = -.352$, $p = .026$, with an $R^2 = .124$. All other correlations were not significant between the self-reported sleep quality questionnaire and the objective measures.

There was a negative correlation between workload levels and risk index (RI) scores after driving $r(40) = -.347$, $p = .028$, with an $R^2 = .120$. All other correlations were not significant between workload levels and all the other objective measures after driving; m-PVT RTs, 1/RTs, lapses, ATW's DD, ADD, FI, AFI, and ARI scores.

There were no significant correlations between how stressed train drivers felt after driving and all the other objective measures after driving; m-PVT RTs, 1/RTs, lapses, ATW's DD, ADD, FI, RI, AFI, and ARI scores.

There was a significant negative correlation between how much effort train drivers had to put into their job and m-PVT mean number of lapses $r(40) = -.358$, $p = .023$, with an $R^2 = .128$. All other correlations were not significant between how much effort train drivers had to put into their job and all the other objective measures after driving; m-PVT RTs, 1/RTs, ATW's DD, ADD, FI, AFI, and ARI scores.

There was a significant negative correlation between how fatiguing the diagram was for train drivers and risk index (RI) scores after driving $r(40) = -.429$, $p = .006$, with an $R^2 = .184$. All other correlations were not significant between how fatiguing the diagram was for train drivers and all the other objective measures after driving; m-PVT RTs, 1/RTs, lapses, ATW's DD, ADD, FI, AFI, and ARI scores.

7.4.9: Discussion

The aim of this study was to investigate whether an offline iOS mobile app version of the online 10-min PVT i.e., offline 10-min m-PVT could be used to detect levels of fatigue in frontline safety critical workers. Firstly, intraclass correlation coefficient (ICC) analysis revealed an excellent degree of reliability in this study (see Ko & Li, 2016, for review) across all 40 mean reaction Times (RTs) which indicates that the offline 10-min m-PVT was able to generate highly reliable RTs across all train drivers. Secondly, these findings also have a higher level of confidence as all significant *p*-values comprised of either a medium or high partial eta-squared (η_p^2) effect size.

7.4.9.1: Objective measures

7.4.9.1.1 offline 10-min m-PVT

This study generated reaction times (RTs) using the offline 10-min m-PVT that were comparable to those found in the literature (Basner *et al*., 2011; Basner & Dinges, 2011; Dinges *et al*., 1987; Dinges & Powell, 1985; Dorrian *et al*., 2007; Evans *et al*., 2019; Khitrov *et al*., 2014; Lamond *et al*., 2005; Lamond *et al*., 2008; Loh *et al*., 2004; Roach *et al*., 2006) as well as those found in the previous two chapters. In this study it was found that train drivers were 2.96 per cent slower in their RTs after driving when compared to before driving for the high workload day. This is further reinforced by the speed response (1/RTs) outcome metrics, which also revealed that train drivers were 3 per cent worse in their 1/RTs after driving when compared to before driving for the high workload day. In contrast, for the low workload day, there were no significant differences found in train drivers RTs and 1/RTs. In other words, findings from this study seem to indicate that for the high workload day, train drivers were significantly more fatigued after finishing work when compared to before driving. However, for the low workload day, there were not significant differences found. These findings support Neville *et al*. (1994) who identified that extended work periods are associated with higher fatigue levels. Therefore, these findings reinforce the notion that prolonged working hours and displaced shiftwork results in physiological and mental fatigue (Mallis *et al*., 2004).

Findings from this study also support previous research that have identified an increase in fatigue results in impaired alertness (Dorrian *et al.*, 2000; Evans, *et a*l., 2019; Gillberg *et al.*, 1994), whereby the time-on-task effect (Langner & Eickhoff, 2013; Mackworth, 1948; Mackworth, 1968), as measured in mean RTs and mean 1/RTs significantly differed after 5 minutes of continuous performance using the offline 10-min m-PVT. As a result, these findings further support previous work, which suggested that sustained attention drops with prolonged duration of task (Dinges & Powell, 1988; Dinges & Powell, 1989; Doran *et al*., 2001; Evans, *et a*l., 2019). These findings further support previous studies that have also been able to demonstrate that a shorter version of the 'gold standard' 10-min PVT can exhibit the time-on-task effect after only 5 minutes on the PVT (Evans, *et a*l., 2019; Roach *et al*., 2006). However, sustained attention in this study was not found to be as low as 3 minutes as identified by Basner *et al*. (2011) using the PVT brief version (PVT-B). Nevertheless, the PVT-B did differ in their random inter-stimulus intervals (ISI), which was between 1 and 4 seconds, and thus significantly shorter than the standard and accepted ISI of between 2 and 10 seconds. Analyses from the number of lapses did not reveal any additional information as all main effects and interactions were not significant. This was not too surprising when considering that train drivers are rigorously screened and selected based on various desirable criteria, such as higher than average vigilance and alertness levels. According to the RSSB (2015f), the rail industry standard for train driver selection includes the ability to have a quick and adequate response to simple and complex visual and acoustic stimuli.

7.4.9.1.2 Fatigue Risk Index (FRI) calculator

Findings from this study also identified that train drivers fatigue index (FI) scores did not significantly differ between the high workload day when compared to the low workload day, even after adjusting the fatigue index (AFI) for the indiscretions in the recorded scores within Arriva Train Wales' (ATW) CrewPlan scheduling. These findings were an unexpected surprise, as a significant difference in the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator's biomathematical model (BMM) was expected for both the high workload day and low workload day. These findings further support findings from Chapter 3 that identified that the current objective measure of fatigue used at Arriva Trains Wales (ATW) may not be an effective or accurate predictor of train driver's FI scores. However, a statistical difference was found in train drivers recorded risk index (RI) scores between the high workload day when compared to the low workload day as well as a statistical difference in the adjusted risk index (ARI) scores between the high workload day when compared to the low workload day. These findings seem to indicate that perhaps the generated RI scores might be a better predictor of train driver's fatigue levels than the actual generated FI scores itself.

7.4.9.1.3 Diagram duration (DD) and adjusted diagram duration (ADD)

Analyses of the diagram duration (DD) as well as the adjusted diagram duration (ADD) revealed that train drivers seemed to have selected perceived high workload and perceived low workload based on the actual DD for the high workload (i.e., 9 hours 0 minutes) was significantly longer in duration when compared to the low workload (i.e., 7 hours 47 minutes). This was also true after ADD, which also revealed that the high workload (i.e., 8 hours 44 minutes) was significantly longer in duration when compared to the low workload (i.e., 6 hours 57 minutes). These findings seem to indicate that train drivers were identifying and selecting diagrams that were significantly longer in duration for the high workload day when compared to the low workload day – even post adjustment for those train drivers being relieved early of duty e.g., spare (SP), standby (SB), team briefing (TB), etc. at the discretion of the resource management team. These findings are consistent with Fan and Smith (2017a) who have identified that at Arriva Train Wales (ATW), high workload levels resulted in higher self-reported levels of fatigue, as well as workload being a contributing factor towards fatigue (Fan & Smith, 2019). As a result, thorough examination and evaluation of the various high workload diagrams and its impact on train drivers fatigue levels need to be carefully explored in order to reduce workload, which has been extensively linked to increased fatigue levels (Dorrian, Baulk, & Dawson, 2011; Goel *et al.*, 2014; Honn *et al.*, 2016; Kathner *et al.*, 2014; Volker *et al.*, 2016), since increased workload levels have been shown to have significant implications on safety incidents (Edworthy *at al*., 2018; Thomas *et al*., 2019).

7.4.9.2: Subjective single-item measures

7.4.9.2.1 Before driving single-item measures

In the present study it was found that when comparing train drivers responses before driving between the high workload day and the low workload day, '*how alert do you feel now?*' was the only single-item measure that was found to be significantly higher for the high workload day when compared to the low workload day. All other single-item measures before driving i.e., '*how was the quality of your sleep last night?*', '*how was the quality of your sleep the night before last night?*', and '*how well are you feeling now?*' were found not to be significantly different when comparing train drivers responses for the high workload day with the low workload day. These findings are consistent with Dorrian *et al*. (2000) and Gillberg *et al*. (1994) who have identified that an increase in fatigue results in impaired alertness. In the present study it was also found that all train drivers' responses before driving showed positive correlations with each other for both the high workload day and low workload day, independently. These findings seem to indicate that sleep quality, whether the night prior or the previous night prior to the shift were positively associated with both how well train drivers were feeling before driving and how alert train drivers were feeling before driving. These results support Kelley, Feltman and Curry's (2018) findings who have identified that insufficient rest, and poor sleep quality are factors that contribute towards fatigue and performance degradation. This is further supported by Tononi and Cirelli (2006) who argued that cognitive impaired performances are the most rapidly occurring consequence of the effect of sleep deprivation. Therefore, in order for train drivers to be able to better mitigate the ill effects of fatigue, sufficient sleep is a vital component that must be secured – especially from shift workers (Gorlova *et al*., 2019).

7.4.9.2.2 After driving single-item measures

In the present study it was found that when comparing train drivers' responses after driving between the high workload day and the low workload day, it was found that all train drivers' responses after driving for the high workload day were significantly higher when compared to the low workload day. In other words, in this study it was found that train drivers reported; feeling more fatigued, had higher workload, felt more stressed, had to put more effort into their job, and found their diagram to be more fatiguing for the high workload day when compared to responses for the low workload day. These findings are consistent with those found in the literature, which have shown that shift length significantly increases fatigue levels, workload levels, stress levels, and amount of effort required to be put into the job (Dorrian, Baulk, & Dawson, 2011; Guo, Wang, & Ning, 2017; MacDonald & Bendak, 2000; Patterson *et al*., 2012; Patterson *et al*., 2019; Tucker, Smith, Macdonald, & Folkard, 1998). This is further supported by the diagram duration (DD) and adjusted diagram duration (ADD) analyses, which identified that one of the major differences between the high workload day and the low workload day was found to be shift length (i.e., diagram duration), especially since the diagram duration of both the DD and ADD for the high workload day were on average 13.5 per cent (i.e., 1 hour 13 minutes) and 20.4 per cent (i.e., 1 hour 47 minutes) respectively, longer than the low workload day. Therefore, it was not too surprising that findings from this study seemed to indicate that the diagrams that train drivers selected to reflect their high workload day were found to be significantly more fatiguing when compared to the diagrams train drivers selected to reflect their low workload day.

An examination of train drivers' responses after driving from the high workload day revealed that fatigue levels were positively associated with; workload, stress, how much effort train drivers had to put into their job, and how fatiguing the diagram was for train drivers. Conversely, train drivers' responses after driving from the low workload day revealed that fatigue levels were also positively associated with; workload, stress, and how fatiguing the diagram was for train drivers, but not with how much effort train drivers had to put into their job. These findings are consistent with other studies that have found relationships between; fatigue and workload (e.g., Baulk *et al*., 2007; Grech, Neal, Yeo, Humphreys, & Smith, 2009; Gui *et al*., 2015; Kathner *et al*., 2014; Motamedzade *et al*., 2017; Remmen, Herttua, Riss-Jepsen, & Berg-Beckhoff, 2017) and fatigue and stress (e.g., Doerr *et al*., 2015; Fan & Smith, 2017a; MacDonald, 2003). However, only for the high workload day were fatigue levels found to have a positive association with how much effort train drivers had to put into their job. As a result, only responses from the high workload day seemed to support other studies from the literature that have found relationships between fatigue and effort (Dragone, 2009; Fukuda *et al*., 2010; Iodice *et al*., 2017).

In addition, it was also found that workload levels were positively associated with how much effort train drivers had to put into their job for both the high workload day as well as for the low workload day. This finding supports several studies that have also found associations between workload and effort (see Honn *et al*., 2016; Moore *et al*., 2015; Shaw *et al*., 2018, for reviews). There was also a positive association between train drivers reported workload levels and how fatiguing the diagram was for them for both the high workload day as well as for the low workload day. These findings were not too surprising as it has been extensively identified in the literature that workload increases fatigue levels as well as that workload is associated with fatigue levels. For example, Baulk *et al*. (2007) examined the relationship between workload and fatigue and found that fatigue increased with higher workload. This is further supported by Kathner *et al*. (2014) who have identified that high workload levels increase fatigue levels. Conversely, Motamedzade *et al*. (2017) found that mental workload and fatigue are associated with each other. Therefore, Gui *et al*. (2015) state that after continuous and prolonged mental workload, self-reported levels of fatigue increases. However, Grech *et al*. (2009) found that the relationship between workload and fatigue changes over consecutive days, which were not explored in the present study due to the potential impact of prolonged participation on operational safety as well as the practical and logistical access to frontline safety critical train drivers beyond the already 80 minutes commitment.

There was also a positive association between stress and how fatiguing the diagram was for train drivers (i.e., job demand) but only for the high workload day. This result is consistent with Leung, Bowen, Liang and Famakin (2015) who found that job demand predicts stress levels. This finding is further supported by the fact that there was no association between stress levels and job demand for the low workload day. However, research carried out by Wong, Lin, Liu and Wan (2014) in frontline safety critical firefighters found that stress levels was fully mediated by work/leisure conflict and work/family conflict. However, Affrunti, Mehta, Rusch and Frazier (2018) state that socio-economic opportunities and health opportunities influence the association between stress and job demands. Therefore, it seems that perhaps there are additional factors that could explain this association. However, these were not explored in the present study due to accessibility and time constraints of frontline safety critical operational train drivers.

A positive association was only found between stress levels and how much effort train drivers had to put into their job for the low workload day but not the high workload day, despite the fact that both stress levels and how much effort train drivers had to put into their job were found to be significantly higher on the high workload day when compared to the low workload day. This result is not constant with several other studies who have examined the relationship between workload and stress and found that self-reported stress levels during high workload were significantly higher than self-reported stress levels during low workload (Prytz & Scerbo, 2015; Searle, Bright, & Bochner, 2001; van der Meij, Gubbels, Schaveling, Almela, & van Vugt, 2018). Roscoe (1978) identified that workload-related factors significantly increased stress levels. Conversely, van den Hombergh *et al*. (2009) argued that job stress due to workload significantly impairs performance. However, perhaps since some train drivers selected spare or standby for their low workload day – the premise of simply sitting down in the break room waiting to be assigned a diagram or partial diagram that may or may not happen could result in elevated anticipation levels, which could increase stress levels. Neubauer, Smyth and Sliwinski (2018) outlined that the anticipation of a stressor in itself is comparable with the actual experience of the stressor. In contrast, Feldman and Hayes (2005) argue that anticipating a stressor can provide an individual with the time required to be able to apply decision-making and problem-solving techniques as an effective coping strategy.

There was also a positive association between how much effort train drivers had to put into their job and how fatiguing the diagram was for train drivers for both high workload day and low workload day. These findings were in line with the notion that fatigue occurs when effort is exerted beyond worker's limited amount of renewable resources (Dragone, 2009). However, Tucker (2003) argues that fatigue can be recovered through effective periods of rest-breaks to off-set the negative effects of shift durations.

7.4.9.3: Relationships between before driving and after driving single-item measures

When examining the relationships between before driving and after driving, it was found that train drivers' quality of sleep the night before stating their shift was negatively associated with fatigue levels after driving for both the high workload day and low workload day. These findings support RSSB (2003) who have identified that when hours of sleep are missed by train drivers, the need for sleep i.e., 'sleep debt' will significantly increase, which results in both a decrease in alertness and performance (McGuffog, Spencer, Stone, & Turner, 2005; McGuffog, Spencer, Turner, & Stone, 2004; Turner & Stone, 2004). There were also positive associations found between the quality of sleep the previous night prior to train drivers stating their shift and workload levels as well as how much effort train drivers had to put into their job, but only for the high workload day and not the low workload day. These findings further support the notion that sleep debt does not just impair alertness and performance but may also impact train drivers self-reported workload levels as well as the amount of effort they have to be put into their job. Therefore, Zee and Goldstein (2010) recommend that good sleep practices, as well as the application of circadian principles for shift workers could be applied to significantly improve sleep quality, alertness, performance, and safety.

There was a negative association between how well train drivers felt before driving and how stressed train drivers felt after driving for both the high workload day and low workload day. These findings seem to indicate that the higher train drivers' wellness levels were before driving, the less stressed train drivers felt after driving, regardless of their diagram intensity. Jobin, Wrosch and Scheier (2014) have highlighted that optimism can buffer the association between perceived stress levels and raised levels of diurnal cortisol when individuals have higher-than-normal stress levels. Conversely, Cabras and Mondo (2018) argue that optimism and coping strategies strongly influence life satisfaction. However, Cushway and Tyler (1994) have identified that the factors most likely to alleviate stress were better support from colleagues and management.

In addition, there was also a negative association for the high workload day between how alert train drivers felt before driving and how stressed train drivers felt after driving. This finding seems to indicate that the more alert train drivers felt before driving on a high workload day, the less stressed they felt after driving. This finding is consistent with

Leproult, Copinschi, Buxton and VanCauter (1997) who had found that sleep loss results in elevated cortisol levels.

There was also a negative association for the low workload day between how well train drivers felt before driving and how fatigued they felt after driving. This finding seems to indicate that the higher train drivers' wellness levels were before driving on the low workload day, the less fatigued they felt after driving. In addition, there was also a negative association for the low workload day between how stressed train drivers felt before driving and how fatigued they felt after driving. Findings from this study also seem to indicate that the more alert train drivers felt before driving, the less fatigued they felt after driving. These findings further support the need for train drivers to be able to develop effective coping strategies. Lian *et al*. (2016) state that establishing effective coping strategies can significantly reduce the impact of shiftwork. As a result, shift workers utilise various countermeasures as well as behavioural coping strategies to be able to maintain both safety and performance in the workplace (Åkerstedt & Landström, 1998).

7.4.9.4: Caffeine consumption

In this study it was found that train drivers consumed significantly more caffeine by the end of the high workload day when compared to the end of the low workload day. These findings support the view that caffeine consumption is commonly used to improve alertness during the working day (Akerstedt & Ficca, 1997; Smith, 2005), as well as caffeine consumption being found to improve reaction times (Einother & Giesbrecht, 2013; Nehlig, 2010, Smith, 2002). Therefore, Smith (2005) outline that caffeine consumption did not just increase alertness levels but was also associated with fewer cognitive failures and accidents at work. As a result, it has been argued that caffeine consumption may be a positive shiftwork countermeasure (Walsh *et al*., 1995) as it has been found to improve performance as well as positive moods (Sutherland & Christopher, 2005). Therefore, caffeine may be an effective coping strategy for sustaining attention, alertness and positive mood when performance and alertness levels are depilating or significantly low (Brice & Smith, 2001). However, it is important to once again reiterate that there was a significant difference in the average diagram duration (DD) and adjusted diagram duration (ADD) between the high workload day and the low workload days. Therefore, it is probable that the increase in caffeine consumption seen may be due to the significantly longer DD and ADD rather than train drivers simply consuming more caffeine as a coping strategy to retain alertness levels for safety.

For the high workload day, there was a positive association found between accumulated caffeine consumption after driving and workload levels. This finding seems to indicate that as the half-life effect of caffeine diminishes, train drivers' true fatigue level emerges. Various studies have report elimination half-lives of caffeine (Grant, Magruder, & Friedman, 2018) ranging from 3 hours – 10 hours (Parsons & Neims, 1978; Rizzo, Stamps, & Fehr, 1988; Robertson, Wade, Workman, Woosley, & Oates, 1981). Therefore, since train drivers were found to consume significantly more caffeine during the high workload day when compared to the low workload day as well as a positive association between accumulated caffeine consumption after driving and perceived workload levels, it seems that perhaps there may be a trade-off between the desired optimal alertness level and caffeine consumption threshold needed but without compromising as much as possible the impact of caffeine on sleep quality. McLellan *et al.* (2016) highlighted that under normal day-to-day circumstances, there is evidence to suggest that caffeine consumption is modulated until a self-perceived optimal peak level of arousal and cognitive performance is achieved (Harvanko *et al.*, 2015). In contrast, there was a negative correlation between caffeine consumption after driving on the low workload day and how fatiguing the diagram was for train drivers. This finding is consistent with Snel and Lorist

(2011) who pointed out that caffeine enhances various aspects of mental and cognitive function.

7.4.9.5: Relationships between objective and subjective single-item measures

Further analysis revealed a positive association for the high workload day between train drivers subjective single-item measure of how fatigued they felt after driving and the RI scores from the HSE's FRI calculator. In addition, there was a positive association between train drivers' workload levels after driving and the RI scores from the HSE's FRI calculator. This finding further supports the notion that perhaps the RI scores of the HSE's FRI calculator may be a better predictor of fatigue than the FI scores itself. In addition, there were also positive associations for the high workload day between how fatiguing the diagram was for train drivers after driving and all HSE's FRI calculator generated scores i.e., FI, RI, as well as the adjusted scored i.e., AFI, and ARI scores. These findings seem to indicate that the more fatiguing the diagram was reported to be after driving, the higher the generated scores within the HSE's FRI calculator. This result further supports the notion that train drivers selected diagrams for the high workload days predominantly based on the diagram duration. Therefore, since the HSE's FRI calculator marginally increases with cumulative shift duration, these findings were not too surprising.

There was also a negative association between how fatiguing the diagram was for train drivers after driving and their reaction times (RTs) as recorded by the offline 10-min m-PVT but only for the high workload day as there were no associations found for the low workload day. This finding seems to indicate that the more fatiguing the diagram was reported to be after driving for the high workload day, the smaller the difference in their RTs after driving (i.e., difference in RTs between 1-min and 10-min), which suggests potentially slower reactions. Moreover, there was also a positive association for the high workload day between how fatiguing the diagram was for train drivers after driving and their speed responses (1/RTs) also recorded by the offline 10-min m-PVT, but also only for the high workload day as there were no associations found for the low workload day. This finding seems to indicate that the more fatiguing the diagram was reported to be after driving, the greater their 1/RTs after driving (i.e., difference in RTs between 1-min and 10-min), which suggests potentially higher 1/RTs. These findings are consistent with other studies that have found moderate to strong correlations between subjective alertness and performance (e.g. Dorrian *et al*., 2000; Dorrian *et al*., 2003; Gillberg, Kecklund, & Akerstedt, 1994). However, Leproult *et al.* (2003) found that sleep deprivation has independent brain mechanisms that control both subjective and objective alertness. In addition, research carried out by Zhou *et al*. (2012) found that subjective feelings of alertness do not accurately reflect performance on tasks requiring sustained attention. Therefore, caution should be taken to ensure that train drivers self-reported measures and the offline 10-min m-PVT are not used in isolation. For example, it is possible that train drivers, knowing they would potentially have a low workload day, would perhaps go to bed a little later than usual. In contrast, it is also possible that train drivers knowing they would have a high workload day would ensure they have sufficient and adequate rest in preparation for their high workload day. In other words, it seems that train drivers are taking the necessary steps to ensure they are 'fit for duty'.

In the present study there was no association found between how fatigued train drivers felt after driving and all offline 10-min m-PVT outcome metrics. These results were consistent with Lim *et al*. (2010), who found no significant associations between self-reported mental fatigue changes and objective performance decline using the psychomotor vigilance task (PVT). Moreover, these were also similar to those found by Leproult *et al.* (2003) and van Dongen, Maislin, Mullington and Dinges (2003), who also found no significant association between subjective and objective measures of alertness in a group of sleep-deprived participants. Therefore, findings from this study further supports Lim *et al.*'s (2010) argument that participants may not be fully or consciously aware of the degree of their objective impairment after a period of high cognitive workload. These findings are consistent with those found by Honn *et al.* (2016) who identified that in frontline safety critical air pilots, an increase in multi-segment duty with multiple take-offs and landings (i.e., high workload) significantly increased fatigue levels when compared to a duty day of equal duration but with only a single take-off and landing (i.e., low workload). It is important to state that this study was not without limitations.

7.4.9.6: Limitations

Since train drivers were instructed to select both a high workload day and a low workload day, which did not have to be on consecutive days over the coming weeks/months based on their upcoming scheduled diagrams as well as individual availability and practicality, cumulative fatigue was a major limitation that could have potentially impacted the findings of the present study. According to several researchers, cumulative fatigue (see Anderson, Grunstein, & Rajaratnam, 2013; Carskadon & Dement, 1981; Dinges *et al.*, 1997; Folkard & Lombardi, 2006; Landrigan *et al.*, 2014; Rajaratnam & Jones, 2004, for reviews) has been demonstrated to impair performance as well as increase safety incidents (e.g., Chang & Ju, 2008; Grech *et al.*, 2009; Spencer *et al.*, 2006). As a result, there was a clear need to determine whether there were any differences in the number of days train drivers worked prior to taking part in both the high workload day and low workload day. In this study it was found that there were no significant differences in the number of days train drivers worked prior to taking part in both the high workload day and low workload day.

Another limitation of the present study is that even though the offline 10-min m-PVT was developed into an iOS app that could be downloaded and installed by any train driver with an Apple iPhone – the researcher decided to provide train drivers with an Apple iPhone, which were all factory reset manually to ensure that both hardware (i.e., Apple iPhone 6s Plus) and software (i.e., iOS version 10.3.2) configurations were as identical as possible to one another. This was due to the fact that previous research has identified that different mobile devices generate different behavioural observed reaction times (see Evans *et al*., 2019, for review). However, by simply using an Apple iPhone 6s Plus, the researcher inveterately introduced a controlled laboratory environment constraint, which is not realistic or representative of the cohort since not all train drivers have Apple iPhone mobile devices.

Another limitation of the present study was that single-item measures were used to record train drivers self-reported sleep quality and workload. Fisher, Matthews and Gibbons (2016) argue that there have been concerns over the use of single-item measures to assess psychological constructs. These concerns have highlighted that single-item measures do not have content validity due to the fact that single items have criterion deficiency as well as viewed as unreliable since internal consistency estimates of reliability cannot be calculated (Fisher *et al*., 2016; Nagy, 2002; Schriesheim *et al*., 1991). However, recent research has demonstrated extensive evidence that single-item measures are able to provide reliable and valid measures of well-being related constructs (Williams & Smith, 2018). In addition, Fisher *et al*. (2016) have counterargued that even though multiple item scales are more likely to have superior psychometric properties, there are several compelling justifications for why researchers might consider the use of single-item measures (e.g., minimising respondent burden, reducing criterion contamination, increasing face validity). There is also extensive evidence in the literature to support the notion that single-item measures have been successfully administered to investigate; sleep quality (Hughes, Ulmer, Gierisch, & Howard, 2018; Snyder, Cai, DeMuro, Morrison, & Ball, 2018), fatigue (Kim & Abraham, 2017; van Hooff, Geurts, Kompier, & Taris, 2007), mental workload (Monfort, Graybeal, Harwood, McKnight, & Shaw, 2018), stress (Littman, White, Satia, Bowen, & Kristal, 2006), quality of life (de Boer et al., 2004), as well as health (Bowling, 2005).

7.5: Chapter summary

The aim of this chapter was to investigate whether an offline iOS mobile app version of the online 10-min PVT i.e., offline 10-min m-PVT could be used to detect levels of fatigue in frontline safety critical workers. A total of 40 train drivers (35 male and 5 female) mean age 43 years (SD = 7.4), which represented 6.4 per cent of all drivers across the franchise were voluntarily recruited to take part in this study. The study involved train drivers attending four sessions, each lasting approximately 20-minutes; a before driving session (i.e., before booking-on) and an after driving session (i.e., after booking-off) on two different days; one day that train drivers perceived to be a high workload day and another day that train drivers perceived to be a low workload day, in exchange for £20. In this study it was found that train drivers were significantly more fatigued as indicated by slower reaction times (RTs) and higher speed responses (1/RT) on the offline 10-min m-PVT temporal attention task after driving when compared to before driving on the high workload day. In contrast, train drivers were not significantly more fatigued after driving when compared to before driving on the low workload day. However, it was also found that train drivers were more fatigued before driving on the low workload day when compared to the high workload day. The study also found a significant main effect of time on task. Post-hoc tests showed that participants RTs and 1/RTs were significantly slower and greater, respectively, after only 5 minutes on the m-PVT temporal attention task. These findings seem to further validate that an offline iOS mobile app version of the ‘gold standard’ 10-min PVT could be used to detect sensitivity levels of fatigue even after less than 5 minutes on the task for frontline safety critical train drivers as well as for other applied safety critical workers such as hospital staff, emergency services, law enforcement, etc. However, there was no association found between how fatigued train drivers felt after driving and all offline 10-min m-PVT outcome metrics. Furthermore, despite there being a significant difference in the diagram duration (DD) and adjusted diagram duration (ADD) for the high workload day when compared to the low workload day, only the risk index (RI) scores and not the fatigue index (FI) scores were found to be significantly higher for the high workload day when compared to the low workload day. These results were also consistent even after adjusting for indiscretions i.e., adjusted fatigue index (AFI) and adjusted risk index (ARI) in the generated scores from the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator within Arriva Train Wales' (ATW) CrewPlan scheduling. These findings seem to indicate that perhaps RI scores might be a better predictor of train driver's fatigue levels than the FI scores. Correlations were also found which further support this notion, since for the high workload day there was a positive association between train drivers fatigue levels after driving and the RI scores from the HSE's FRI calculator, as well as a positive association between train drivers workload levels after driving and the RI scores from the HSE's FRI calculator. In this study it was found that train drivers consumed significantly more caffeine by the end of the high workload day when compared to the end of the low workload day. However, these findings could be due to the fact that for the high workload day the average DD and ADD were on average significantly longer 13.5 per cent and 20.4 per cent, respectively, when compared to the low workload day. Therefore, it is probable that the increase in caffeine consumption seen may be due to the shift duration rather than train drivers potentially consuming more caffeine as a coping strategy to retain alertness levels for the purpose of operational safety. Further research is now needed to investigate whether a shorter offline iOS mobile app version (i.e., 5-min m-PVT) could still be sensitive enough at detecting levels of fatigue in frontline safety critical settings in train drivers and beyond, e.g., hospital staff, emergency services, law enforcement, etc.

Chapter 8: General Discussion

8.1: Overview of chapter

Chapter 8 presents an integrated discussion of the research described in this thesis. Foremost, the chapter provides an overview of the research undertaken. This is followed by a summary of the main empirical findings and in-cab observations in relation to the objectives set out in the overarching thesis. This discussion will then lead to the main research limitations, followed by practical recommendations as well as recommendations for future research. Finally, this thesis will present recommendations going forward for Transport for Wales Rail Services.

8.2: Overview of research

The overarching aim of the thesis was to develop and validate an objective indicator of fatigue for frontline safety critical workers. This thesis was carried out in partnership with the railway franchise Arriva Trains Wales (ATW). The rationale for developing an alternative objective indicator of fatigue for frontline safety critical workers was due to the fact that current biomathematical model (BMM) of fatigue used at ATW may not be an effective or accurate predictor of train drivers' fatigue levels.

In the first objective, analyses of all 578 accessible safety incidents in which fatigue could have been a contributing factor i.e., Signal Passed at Danger (SPAD), Train Protection & Warning System (TPWS) activation, Automatic Warning System (AWS) slow to cancel, failed to call, and station overrun – only identified 0.2 per cent and 2.1 per cent of fatigue index (FI) scores and risk index (RI) scores, respectively that exceeded ATW's Fatigue Risk Management System (FRMS) BMM recommended thresholds.

In the second objective, in-cab observations from within the cab environment identified that; cab noise exposure (whether internal or external), cab temperature, partial unit improvements, and cab working conditions were major issues that could further contribute towards safety incidents when fatigued. As a result, further research was needed to determine whether an alternative objective indicator of fatigue could be developed and validated to support the FRMS at ATW.

In a controlled laboratory setting, the Psychomotor Vigilance Task (PVT) has become the widely accepted 'gold standard' tool for assessing the impact of sleep deprivation and fatigue on human cognitive neurobehavioral performance for monitoring temporal dynamic changes in attention. Several empirical studies were carried out to replicate and validate an alternative online mobile version of the 10-min PVT (see Evans *et al*., 2019, for review) i.e., online 10-min m-PVT (third objective), a shorter version i.e., online 5-min m-PVT (forth objective) as well as developing an offline iOS mobile app version i.e., offline 10-min m-PVT (fifth objective) for frontline safety critical workers. Firstly, intraclass correlation coefficient (ICC) analyses of all three alternative mobile versions of the 'gold standard' 10-min PVT revealed excellent degree of reliability (see Ko & Li, 2016, for review) .960 (online 10-min m-PVT), .970 (online 5-min m-PVT), and .990 (offline 10-min m-PVT), which seems to indicate that a mobile version of the 'gold standard' 10-min PVT has test-retest and intrarater reliability and validity.

Findings from these studies identified that the online 10-min m-PVT was sufficiently sensitive at detecting levels of fatigue between the morning and afternoon (i.e., time-of-day effect) as well as sensitive enough at detecting levels of fatigue after 4 minutes on the task, while the shorter version i.e., online 5-min m-PVT was also sufficiently sensitive at detecting levels of fatigue in simulated workload (i.e., high simulated workload and low simulated workload). However, an interaction between simulated workload and time-on-task revealed that caution needs to be taken when administering a shorter version of the 'gold standard' 10-min PVT i.e., online 5-min m-PVT as only the low simulated workload condition found a significant time-on-task effect while no significant differences were found for the high simulated workload condition. The offline 10-min m-PVT was administered using frontline safety critical train drivers and it was found that the offline 10-min m-PVT was sensitive enough at detecting levels of fatigue in high workload diagrams. Further research is now needed to investigate whether a shorter offline iOS mobile app version i.e., offline 5-min m-PVT could still be sensitive enough at detecting levels of fatigue in high workload diagrams for frontline safety critical workers such as train drivers, hospital staff, emergency services, law enforcers, etc.

8.3: Summary of the main research findings

8.3.1: Objective 1

The aim of the first objective was to investigates the effectiveness of the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator used at Arriva Trains Wales (ATW) as part of their Fatigue Risk Management System (FRMS) for monitoring and managing safety incidents in which fatigue could have been a contributing factor. This objective set out to answer two fundamental questions:

1. The aim of the first study was to investigate whether the present biomathematical model (BMM) for assessing train drivers' level of fatigue at Arriva Trains Wales (ATW) i.e., the Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator related to the number of safety incidents in which fatigue could have been a contributing factor.

2. The aim of the second study was to investigate whether restricting train drivers from working on their assigned rest days based on a naturally occurring intervention reduced the number of safety incidents in which fatigue could have been a contributing factor.

8.3.1.1: Study 1

The aim of the first study was to investigate whether restricting train drivers from working on their assigned rest days reduced the number of safety incidents in which fatigue could have been a contributing factor. In this study it was found that the current Health and Safety Executive's (HSE) Fatigue Risk Index (FRI) calculator for assessing train drivers' level of fatigue used at Arriva Trains Wales (ATW) may not be an effective or accurate predictor of train drivers' fatigue or risk levels, since analyses of all 578 accessible safety incidents in which fatigue could have been a contributing factor only identified 0.2 per cent and 2.1 per cent of FI scores and RI scores respectively, that exceeded ATW's Fatigue Risk Management System (FRMS) recommended thresholds of FI score = 45 and RI score = 1.6 (ATW, 2017a). Therefore, further analysis was needed to investigate whether restricting train drivers from working on their assigned rest days based on a naturally occurring intervention reduced the number of safety incidents in which fatigue could have been a contributing factor.

8.3.1.2: Study 2

The aim of the second study was to investigate whether restricting train drivers from working on their assigned rest days based on a naturally occurring intervention reduced the number of safety incidents in which fatigue could have been a contributing factor. In this study, the observed frequency of safety incidents during the naturally occurring intervention time period was compared across the same dates in previous years to be able to determine whether restricting train drivers from working their assigned rest days significantly reduced the number of safety incidents in which fatigue could have been a contributing factor. In this study it was found that there were no statistical differences in safety incident frequencies throughout the six time periods. However, there was a 47 per cent statistically significant decrease in the observed frequency of safety incidents between the time period of the naturally occurring intervention (i.e., 2016) and the preceding year (i.e., 2015). These findings seem to indicate that perhaps other external influencers such as individual and environmental factors may also be contributors, which have not been incorporated into the biomathematical model within the Health and Safety Executive (HSE) Fatigue Risk Index (FRI) calculator. As a result, there is a clear need for better understanding some of the environmental factors that could contribute towards safety incidents in frontline safety critical workers.

8.3.2: Objective 2

The aim of the second objective was to identify some of the external environmental factors that could contribute towards safety incidents when fatigued. Through the process of ethnographic research, the observer was able to complete ~120 hours of consecutive in-cab observations, which allowed the observer to identify that cab noise exposure (whether internal or external), cab temperature, partial unit improvements, and cab working conditions were major observed concerns that could contribute towards safety incidents when fatigued. Therefore, there was a need to validate and develop an alternative objective indicators of fatigue to better support the FRMS at TfWRS as the previous objective has identified that the current HSE's FRI calculator used at ATW may not be an effective predictor of train drivers' fatigue, as well as cab observations highlighting that there are several external environmental factors that the current HSE's FRI calculator does not address or acknowledge that could contribute towards safety incidents when fatigued.

8.3.3: Objective 3

The aim of the third objective was to investigate whether an alternative online mobile version of the 'gold standard' 10-min PVT i.e., online 10-min m-PVT could be used to provide an objective indicator of fatigue for frontline safety critical workers. In this study it was found that there was a significant difference in both reaction time (RT) and speed response (i.e., reciprocal of reaction time) (1/RT) between the morning and afternoon. Post-hoc tests showed that participants RTs were significantly faster in the morning when compared to the afternoon, as well as participants 1/RTs being significantly slower in the morning when compared to the afternoon. This study also found that the online 10-min m-PVT was sufficiently sensitive at detecting levels of fatigue after 4 minutes on the temporal attention task for both RTs and 1/RTs outcome metrics. These findings do seem to indicate that the alternative online 10-min m-PVT could be used to provide an objective indicator of fatigue for frontline safety critical workers, such as; train drivers, hospital staff, emergency services, law enforcement, etc. However, in frontline safety critical settings, the 10-min m-PVT may be too deemed to be too long in duration. Newington and Metcalfe (2014) suggested that in order to improve recruitment, reducing participant burden was essential. Therefore, there was a clear need to investigate whether a shorter mobile version of the online 10-min m-PVT i.e., online 5-min m-PVT could be used to provide an objective indicator of simulated workload fatigue.

8.3.4: Objective 4

The aim of the fourth objective was to investigate whether a shorter mobile version of the online 10-min m-PVT i.e., online 5-min m-PVT could be used to provide an objective indicator of simulated workload fatigue. In this study it was found that for the high simulated workload condition, participants' number of lapses were significantly higher after completing a series of online cognitive tasks designed to elicit mental workload (i.e., high simulated workload) when compared to before completing a series of online cognitive tasks. In contrast, there was no significant difference in participants' number of lapses before and after watching one episode of The Big Bang Theory (i.e., low simulated workload). However, the interaction between simulated workload and time on task revealed that both the RT and 1/RT outcome metrics were only significant for the low simulated workload condition after 5 minutes, with no significant differences found in both the RT and 1/RT outcome metrics for the high simulated workload condition. Therefore, the time on task effect in this study seemed to indicate that caution needs to be taken when administering a shorter version of the ‘gold standard’ 10-min PVT i.e., online 5-min m-PVT as it may not be sensitive or reliable enough to detect levels of fatigue. However, despite studies outlining the availability to run online experiments (Barnhoorn *et al*., 2015; Crump *et al*., 2013; Reimers & Stewart, 2016) and other studies demonstrating comparability between lab based and online based experiments (e.g., de Leeuw & Motz, 2016; Reimers & Stewart, 2007; Schubert *et al*., 2013; Simcox & Fiez, 2014), in rare occasions when the Wi-Fi connectivity dropped, participants’ trials were lost and thus not recorded. As a result, a total of 1.95 per cent of all test trials were never recorded, which despite representing a large reduction of 83.49 per cent Wi-Fi connectivity drop when compared to the previous study, are still relatively high and could potentially impact the validity of the m-PVT. Therefore, there is a clear need for the development of an offline iOS mobile app version of the 10-min m-PVT. Further research is now needed to investigate whether an offline iOS mobile app version of the online 10-min m-PVT i.e., offline 10-min m-PVT could also be used to detect sensitivity levels of fatigue in frontline safety critical workers.

8.3.5: Objective 5

The aim of the fifth objective was to develop and validate an alternative offline iOS mobile app version of the online 10-min m-PVT i.e., offline 10-min m-PVT to detect levels of fatigue for frontline safety critical train drivers. In this study it was found that train drivers were significantly more fatigued as indicated by slower reaction times (RTs) and higher speed responses (1/RT) on the offline 10-min m-PVT temporal attention task after driving when compared to before driving on the high workload day. In contrast, train drivers were not significantly more fatigued after driving when compared to before driving on the low workload day. The study also found that participants’ RTs and 1/RTs were significantly slower and greater, respectively, after 5 minutes on the task. These findings seem to further validate that an offline iOS mobile app version of the 'gold standard' 10-min PVT could be used to detect sensitivity levels of fatigue even after only 5 minutes on the task. Further research is now needed to determine whether a shorter offline iOS mobile app version i.e., 5-min m-PVT could still be sensitive enough at detecting levels of fatigue in frontline safety critical workers such as train drivers and beyond, e.g., hospital staff, emergency services, law enforcers, etc.

8.4: Summary of research limitations

There were several limitations throughout each objective that will now be discussed in great depth. For the first objective, one of the biggest limitations was the issues associated with the access to the various ATW systems required for the researcher to be able to compile the secondary analyses of the large existing data, whereby the various data system had its own unique challenges. For example, in order to be able to extract all the relevant information, the researcher had to firstly manually identify and download all safety incidents using the Safety Management Information System (SMIS) in which fatigue could have been a contributing factor, which were formally identified to be the following five categories; SPAD, TPWS activation, AWS slow to cancel, failed to call, and station overrun. All historical safety incident reports were extracted as far back as the 1st June 2010 when the HSE's FRI calculator was integrated into ATW's CrewPlan scheduling system. A total of 901 safety incident reports in these five categories were identified and downloaded with the relevant information extracted and sorted for later analyses e.g., incident date, incident time, incident group, incident type, region, responsibility operator, driver, depot, location, event logged, etc. The information from the safety incident reports were then used to identify each individual train driver as well as their present working depot, as CrewPlan is far from flexible and requires the end-user to firstly navigate to the correct depot for the train driver in question, followed by identifying the correct train driver e.g., M. S. Evans or M. Evans before being able to generate CrewPlan/s fatigue index (FI) spreadsheet. For each train driver, a representative FI spreadsheet report was downloaded with the relevant information extracted for later analyses e.g., company, driver name, location, commuting time, day, on duty, off duty, diagram turn, shift duration, rest duration, average duty per day, shift type, fatigue index, risk index, mean fatigue, mean risk, etc. Conversely, non-standard reports (e.g., pay number, home depot, actual turn, actual book-on, actual book-off, actual total hours, etc.) and personal roster reports (e.g., pay number, depot, week starting, total hours, book-on, book-off, turn, total, etc.) were also generated and downloaded for verification purposes. This whole process had to be carried out for each train driver, which was immensely time demanding. Furthermore, the whole process required significant levels of concentration as well as introduction of continuous verification checks throughout each step to ensure no errors were introduced. It is also important to acknowledge that despite the whole process requiring significant dedication, only 64.2 per cent of train drivers' FI and RI scores could be accessible from ATW's CrewPlan scheduling system. There were several fundamental factors that resulted in 35.8 per cent of train drivers not being accessible. Poor record keeping was the biggest drawback, for example; in some instances, train drivers were referred to simply as D. Jones. There were multiple D. Jones' across the ATW franchise. Therefore, a process of elimination was constantly implemented. Firstly, checking the depot, if this information was not provided, then checking the potential D. Jones train drivers who worked on the day of the safety incident. In some cases, this was not enough to eliminate all but one train driver. For example, checking the time of the safety incident and validating it with the train driver's scheduled shift e.g., non-standard report or personal roster report. There were a few occasions when these steps were not sufficient. As a result, the train driver could not be identified. Furthermore, there were a large proportion of safety incident reports that did not have the train driver's name. In addition, there were safety incident reports that used the train's head code (i.e., route of the train journey) instead of the train driver's name. For example, CH308 instead of D. Jones or David Jones. Moreover, there were safety incident reports where train driver's names were misspelled, which made it difficult and, in some cases, almost impossible to identify without manually searching through each depot individually. There were also issues with over and under representing safety incident reports within the Safety Management Information System (SMIS). For example, there were instances when a safety incident report was recorded twice in SMIS. Moreover, some safety incident reports omitted the responsible operator. In other words, the responsible operator entry box was not populated. Furthermore, in some cases, under close investigation, safety incidents reports from different responsible operators for example National Rail were wrongly categorised, which meant they were labelled as an ATW safety incident instead of National Rail incident. In addition, FI and RI scores of retired train drivers and train drivers who had left the company were no longer accessible.

For the second objective, the researcher was fortunate enough to be able to carry out ~120 hours cab observations. Beyond the invaluable experience, exposure, and understanding from within the cab environment, one of the major limitations with ethnographic research was that the observer could become biased towards the direction of the research i.e., 'observer effects' (LeCompte & Goetz, 1982). Therefore, despite the researcher having four distinctive train driving instructors; two from Valleys & Cardiff local routes and two from Cardiff mainline routes – the observer was paired with highly experienced and competent instructors, who had clearly developed effective fatigue coping strategies (e.g., Åkerstedt & Landström, 1998; Centofanti *et al.*, 2018; Darwent *et al.*, 2015; Jay *et al.*, 2008; Lian *et al.*, 2016; Richards *et al.*, 2018). Therefore, it is important to acknowledge that these cab observations were also biased in the sense that they may not be a true representation of the train drivers' establishment. As a result, further research is needed for the researcher to be able sit with randomised train drivers with the appropriate clearance access beyond those who are highly experienced and competent i.e., train driving instructors.

One of the biggest limitations with both the online 10-min m-PVT (third objective) and online 5-min m-PVT (forth objective) was that in rare occasions when the Wi-Fi connectivity dropped, participants' trials were never recorded. This accounted for ~8 per cent of all trials for the 10-min m-PVT and ~2 per cent of all test trials for the 5-min m-PVT, which could potentially impact the validity of the m-PVT. Therefore, since the overarching aim of the thesis was to develop and validate an objective indicator of fatigue for frontline safety critical workers, Wi-Fi connectivity outside of a controlled laboratory environment may not always be possible and in some locations, almost impossible. As a result, even though it was identified that the shorter online 5-min m-PVT was sensitive enough at detecting levels of fatigue in simulated workload, these findings were still based from a controlled laboratory setting. Consequently, the offline iOS mobile app version was vital in order to effectively overcome the Wi-Fi connectivity issues of the online version as well as being realistically more practical outside a controlled laboratory environment. It is important to also outline that for both the online 10-min m-PVT and online 5-min m-PVT, participants were instructed not to consume caffeine (e.g., coffee, energy drinks, tea, etc.) and alcohol during the 24 hours before the study, since research has found that higher levels of caffeine consumption resulted in significantly higher alertness levels over the working day (Smith, 2005), significantly reduced reaction times (Einother & Giesbrecht, 2013; Nehlig, 2010; Smith, 2002; Smith, 2005), improved performance (Smith, Sutherland, & Christopher, 2005), and diminishes the true state of fatigue (Grant, Magruder, & Friedman, 2018). However, these restrictions could have significantly impacted the results found. Therefore, taking these factors into consideration, the researcher felt that since the offline iOS mobile app version would be administered on frontline safety critical train drivers while on their normal operational diagrams without restricting caffeine consumption, it was crucial to firstly replicate and validate the offline 10-min m-PVT as it was developed in a completely different programming language to both the online 10-min m-PVT and online 5-min m-PVT. Research carried out by Evans *et al*. (2019) had identified that both hardware and software differences generated slightly different behavioural observed reaction times. As a result, it was more appropriate to administer a 10-minute version rather than a shorter version i.e., 5-minute as the 10-minute version has become the widely accepted 'gold standard' tool for assessing the impact of sleep deprivation and fatigue on human cognitive neurobehavioral performance for monitoring temporal changes in attention (e.g., Belenky *et al*., 2003; Dinges *et al*., 1997; Jewett, Dijk, Kronauer & Dinges, 1999; Lamond *et al*., 2003). However, it is important to further acknowledge that the impact of fatigue on human cognitive neurobehavioral performance can also be influenced by other factors such as individual differences (Leproult *et al*., 2003; Patkai, 1971; Parasuraman *et al*., 2009; RSSB, 2012; van Dongen, 2006) as well as the hypothalamic regulation of sleep and circadian rhythms (Saper, Scammell, & Lu, 2005). There was another limitation that should be acknowledged with the experimental design of the first online 10-min m-PVT study, whereby those findings could also be explained by participants eating and/or drinking before their morning session, since it is highly likely that participants would have had ample opportunities for breakfast before commencing the study at 11:00 when compared to the opportunity to get food prior to the afternoon session i.e., 17:00. However, this limitation was addressed for the second study i.e., the online 5-min m-PVT simulated workload, as participants were all brought in to start the study by 07:30 in the morning.

There were also several limitations with the fifth objective. Firstly, even though the offline 10-min m-PVT was developed into an iOS app that could be downloaded and installed by any train driver with an Apple iPhone – the researcher decided to provide train drivers with Apple iPhone 6s Plus mobile devices, which were all factory reset manually to ensure that both hardware and software configurations were as identical as possible to one another. This was due to the fact that previous research had identified that different mobile devices generate different behavioural observed reaction times (see Evans *et al*., 2019, for review). However, by simply using an Apple iPhone 6s Plus, the researcher inveterately introduced a controlled laboratory constraint, which may not be realistic or representative of the cohort to all have Apple iPhone mobile devices. Therefore, going forward in order to ensure replicability, various mobile models across various manufacturing brands, including different operating system versions should be explored. In addition, the fifth objective also recruited a total of 40 train drivers. Therefore, all associations found using the subjective single-item measures were based on a very limited number of train drivers. As a result, the researcher advices that caution should be taken when examining the various associations found between the objective and subjective measures.

One other limitation to acknowledge from both the simulated workload study and the workload study with train drivers was that even though workload manipulations were carried out, no other variety of psychological and physiological techniques, which include; subjective psychological self-reported measures e.g., the NASA Task Load Index (NASA-TLX) (Byrne *et al.*, 2010; Hart & Staveland, 1988; Orlandi & Brooks, 2018; Shakouri, Ikuma, Aghazadeh, & Nahmens, 2018) and the NASA-MATB (National Aeronautics and Space Administration Multi-Attribute Task Battery (Comstock & Arnegard, 1992) as well as objective physiological measures e.g., heart rate (HR) (e.g., Shakouri, Ikuma, Aghazadeh, & Nahmens, 2018), galvanic skin response (GSR) (e.g., Widyanti, Hanna., Muslim, & Sutalaksana, 2017), body temperature (e.g., Vergara, Moenne-Loccoz, & Maldonado, 2017), electrocardiogram (ECG) (e.g., Heine, Lenis, Reichensperger, Beran, Doessel, & Deml, 2017), electroencephalogram (EEG) (e.g., Berka *et al*., 2007; Hogervorst, Brouwer, & Van Erp, 2014; Hsu, Wang, Chen, & Chen, 2015; Jimenez-Molina, Retamal, & Lira, 2018; Shaw *et al*., 2018; So, Wong, Mak, & Chan, 2017), and eye tracking (Saito, 1992; Xu, Min, & Hu, 2018), which have been extensively examined in various safety critical environments were carried out. These alternative workload measures could have provided a far more rigorous dataset that is richer for comparison with both the online 5-min m-PVT and the offline 10-min m-PVT. However, these additional measures would require significant resources, whether cost driven or time constraints, which is not always feasible when operating in frontline safety critical settings.

There are also limitations that need to be acknowledge when it comes to the population cohort that were recruited in these three m-PVT studies. Firstly, the offline 10-min m-PVT study with train drivers were on average ~23 years older than the previous two studies i.e., the online 10-min m-PVT and online 5-min m-PVT, since the first two studies recruited undergraduate psychology students. Despite Wood *et al*. (2015) identifying that reaction time decreases with age, it is important to outline that train drivers are selected and recruited on the bases of scoring high on pre-defined vigilance, reaction time, concertation levels, and alertness thresholds. Therefore, with an average 10 years' experience since certified as a train driver i.e., holding a Category B train driving licence under the Train Driving Licences and Certificates Regulations 2010 (TDLCR) – these recruited professional train drivers are highly competent. Secondly, male train drivers represented 87.5 per cent of the recruited cohort, while in the previous two studies, males only represented 12.2 per cent and 28.2 per cent, respectively. Historically train drivers in the UK have been predominantly male. To put this into perspective, at Transport for Wales Rail Services (TfWRS), which was previously operated by Arriva Trains Wales (ATW) – female train drivers only represent 4.6 percent of all train drivers (ASLEF, 2019), which is 1.1 per cent below the national average, which currently stands at 6.5 per cent in England, Scotland and Wales (ASLEF, 2019). Despite the third m-PVT study recruiting 5 per cent female train drivers, which accurately reflects the gender inequality among train divers, the previous two studies i.e., online 10-min m-PVT and online 5-min m-PVT had a much larger proportion of female participants.

Another limitation of developing and validating an alternative online and offline mobile version of the 'gold standard' PVT i.e., online 10-min m-PVT, online 5-min m-PVT, and offline 10-min m-PVT was that even though there were multiple additional outcome analyses, only the three most frequently reported PVT outcome analyses in the literature were carried out as identified by Basner and Dinges' (2011) review of 141 journal manuscripts, which were identified as; mean reaction time (RT), mean speed response (1/RT), and mean number of lapses. However, there were other PVT outcome metrics that could have been calculated based on the recorded RT data, such as; the mean median, fastest 10%, slowest 10% but to name a few. The relationship between mean RTs and mean 1/RTs is that mean 1/RTs are the reciprocal of mean RTs latency and is calculated using the following equation $\{[1 \,/\, \text{mean RT}] \times 1000\}$ (Belenky *et al.*, 2003). The relationship between mean RTs and mean number of lapses is that the mean number of lapses is the cumulative mean number of RTs exceeding 500 ms, which has been demonstrated to be a valid indicator of the level of fatigue existing at the time of the test that represents lapses in attention (Belenky *et al.*, 2003; Dinges & Kribbs, 1991).

8.5: Practical recommendations

For the first objective, it was identified that poor record keeping resulted in 35.8 per cent of train drivers' safety incidents not being accessible for analyses. Therefore, it is strongly recommended that Transport for Wales Rail Services (TfWRS) seek consultation for how to best monitor, generate and evaluate key performance indicators of fatigue (see ORR, 2016c; ORR, 2017b, for reviews) in order to better support the implementation of a comprehensive Fatigue Risk Management System (FRMS) at TfWRS. In addition, since analyses of all accessible safety incidents in which fatigue could have been a contributing factor only identified 0.2 per cent FI scores and 2.1 per cent RI scores that exceeded ATW's FRMS recommended thresholds (see ATW, 2017a, for review), TfWRS may need to carefully consider and evaluate with the appropriate urgency whether their current biomathematical model (BMM) i.e., the HSE's FRI calculator is effective at predicting safety incidents in which fatigue could have been a contributing factor in their frontline safety critical workers, especially since ORR (2016a) argue that caution on the FI and RI thresholds should be taken as these are but one component of the FRMS in assessing fatigue, as well as there now being no agreed 'thresholds' for the HSE's FRI calculator (Somvang *et al*., 2016). In addition, it is also recommended that TfWRS start using train drivers' unique payroll number going forward on all safety incident reports accessible via the Safety Management Information System (SMIS) for better identification as well as ensuring anonymity compliance under the General Data Protection Regulation (GDPR). Moreover, it is also strongly recommended going forward that TfWRS carefully evaluate their Safety, Training and Update Day (STUD) programme to better reflect and address the high TPWS Activation safety incidents between 6 – 10 years of train driving experience.

For the second objective, it was fundamentally observed that a large proportion of the various identified external cab environmental factors i.e., noises (whether internal or external to the cab), temperature, and working conditions that could contribute towards safety incidents when fatigued could be reduced or even eliminated completely through effective maintenance or through the introduction of new rolling stock that have been redesigned with these issues in mind. As a result, it is recommended that TfWRS fully considers any new rolling stock that may be added to their current rolling stock fleet incorporates significant soundproofing inside the cab as well as an effective cab climate environment control system that allows train drivers to adjust the temperature based on the train driver's preference and operational necessities e.g., cold air to briefly promote alertness. Moreover, it is also recommended that TfWRS' diagram planning team identifies for each train journey the most frequently recorded delays e.g., overcrowding on commuter services (i.e., delays due to passengers alighting and boarding), wheelchair accessibility, anti-social behaviour, on-board incident, etc. in order to be able to generate diagrams that realistically and proactively reflect operation as well as produce more accurate timetables that meets customers' expectations, including transparency. Furthermore, it is also recommended that TfWRS considers investigating the association between multi-segment operations within the diagram turn (i.e., each unit journey) with train drivers' objective and subjective fatigue levels in order to further improve the current FRMS.

8.6: Recommendations for future research

Further studies are now needed to investigate whether additional components could be integrated into to the current Fatigue Risk Management System (FRMS) at Transport for Wales Rail Services (TfWRS). For example, the ability to generate individual diagram scores based on a pre-defined criterion e.g., duration of diagram, driving duration until first official break, driving duration until second official break, unofficial break e.g., waiting period when changing ends, number of head sections, number of stops per head section, total number of diagram stops, number of signals between stops, speed restrictions per signal section, track condition on signal section, etc., which could then be used to validate, monitor, and continuously review in a manner that reflects changes and incorporates best diagram planning practices.

Since cumulative fatigue (see Anderson, Grunstein, & Rajaratnam, 2013; Carskadon & Dement, 1981; Dinges *et al*., 1997; Folkard & Lombardi, 2006; Landrigan *et al*., 2014;

Rajaratnam & Jones, 2004, for reviews) has been demonstrated to impair performance and increase safety incidents (e.g., Chang & Ju, 2008; Grech *et al*., 2009; Spencer *et al*., 2006), future studies should aim to investigate the effects of train drivers' cumulative fatigue on workload using the offline m-PVT with a significantly larger number of frontline safety critical workers. In addition, further research is also needed to investigate whether a shorter offline iOS mobile app version i.e., 5-min m-PVT could still be sensitive enough at detecting levels of fatigue in frontline safety critical settings in train drivers and beyond, e.g., hospital staff, emergency services, law enforcement, etc. Moreover, future research should attempt to investigate the impact of fatigue on performance for each diagram turn while also controlling for workload by using subjective psychological self-reported measures e.g., the NASA Task Load Index (NASA-TLX) (Byrne *et al.*, 2010; Hart & Staveland, 1988; Orlandi & Brooks, 2018; Shakouri, Ikuma, Aghazadeh, & Nahmens, 2018). Finally, further research is also needed to explore the various effective coping strategies train drivers utilise to mitigate the ill effects of shiftwork and workload.

Future studies could also be carried out across the various depot establishments at TfWRS since this study exclusively only looked at train drivers from two depots i.e., Valleys & Cardiff Local Routes (formally known as Cardiff Valley lines) and Cardiff Mainline, which were both based at Cardiff, Wales. However, TfWRS currently operates across 12 depots split into both the South (i.e. Valleys & Cardiff Local Routes, Cardiff Mainline, Carmarthen, Treherbert, and Rhymney) and the North (i.e., Chester, Crewe, Machynlleth, Shrewsbury, Holyhead, Llandudno Junction, and Pwllheli) regions of the Wales and Borders Franchise. In addition, it would also be advantageous to carry out longitudinal m-PVT studies to look at train drivers RT changes over time.

8.7: The future of train driving

Despite the fact that driving a train is always carried out in complete isolation, Heath *et al*. (1999) stated that due to the high levels of regulations in place, a train does not require a driver. However, a great deal has changed since the late 1990s when automated trains or even automated vehicles were in its infancy. For the last 20 years instead of moving towards automated trains, additional control systems had been implemented within the cab, traction and track e.g., Global System for Mobile Communications – Railway (GSM-R), Driver's Reminder Appliance (DRA), Automatic Warning System (AWS), Train Protection Warning System (TPWS), etc. with the aim to further prevent potentially catastrophic safety critical incidents. Conversely, of late there has been a shift towards automated train systems within the rail industry and due to dramatic advancements in technology, journey length, travel speed, and customer service qualities (Yin *et al*., 2017). Moreover, due to global co-operative climate change policy and action i.e., The Paris Agreement 2015 (Seo, 2017; UNFCCC, 2015) and traffic congestion problems in large cities (Ma, Zhang, & Li, 2016), urban rail transport is now being promoted as the green option as well as the convenient transport method in large cities such as; London, New York, Tokyo, Taiwan, etc. (Wang, Tang, Ning, van den Boom, & De Schutter, 2015).

The Automatic Train Operation (ATO) Programme, which is now part of Digital Railway that involves the close collaboration between various stakeholders e.g., Network Rail, ORR, RSSB, TOCs, etc. to produce ATO control systems on rolling stock, digital signalling and trackside (Network Rail, 2017). Some Train Operating Companies (TOCs) are now in the early stages of integrating the ATO e.g., Thameslink (see Hayat & Redfern, 2018, for review) and Crossrail (see ORR, 2016b, for review), since ATO is considered as an emerging technological option that will eventually replace the traditional manual train driving in many current urban rail networks (Dong, Ning, Cai, & Hou, 2010; Miyatake & Ko, 2010). However, with Thameslink and Crossrail programmes implementing the ATO in 2019, the ability for the ATO to automatically adjust the unit's speed in accordance with the various regulations for aligning to a fixed unit marker within the station platform for the purpose of parking will ensure better timetable planning (Wang, Xiao, Chen, & Li, 2018). As result, more trains can be added to the timetable to reduce crowded platforms as well as to alleviate congested saloons during peak time (Woods, 2012; Woods & Barrett, 2018). However, at present the literature has mainly focused on technical improvements of ATO systems (Caramia *et al*., 2017; Wang *et al*., 2015; Wang *et al*., 2018; Yin *et al*., 2017), with no current peer-reviewed literature on how TOCs will support train drivers in the transition from manually operating the unit to relinquishing control to various automated train systems.

8.8: Concluding remarks

The originality and contribution of this thesis lie in three main domains. Firstly, findings from this thesis identified that the current biomathematical model (BMM) of fatigue i.e., the Health and Safety Executive (HSE) Fatigue Risk Index (FRI) calculator used at Arriva Trains Wales (ATW) in not an effective method for monitoring and reducing safety incidents in which fatigue could have been a contributing factor. Secondly, that ethnographic in-cab observations revealed that noise (whether internal or external to the cab), cab temperature, and cab working conditions were major concerns that could contribute towards safety incidents when fatigued. In addition, it was also found that most of these contributors towards safety incidents when fatigued could be eliminated through the process of replacing the old rolling stock fleet with modernised rolling stock that have addressed these issues. Thirdly, that an alternative objective indicator of fatigue i.e., the offline 10-min m-PVT was sensitive at detecting levels of fatigue due to workload. In summary, this thesis highlights that a mobile app version of the ‘gold standard’ 10-min PVT could be used to further strengthen and complement the current Fatigue Risk Management Systems (FRMS) for frontline safety critical workers.

Analyzing Safety Incidents using the Fatigue Risk Index Calculator as an Indicator of Fatigue within a UK Rail Franchise

Michael Scott Evans, Andrew Paul Smith

***Abstract*—**The feeling of fatigue at work could potentially have devastating consequences. The aim of this study was to investigate whether the well-established objective indicator of fatigue – the Fatigue Risk Index (FRI) calculator used by the rail industry is an effective indicator to the number of safety incidents, in which fatigue could have been a contributing factor. The study received ethics approval from Cardiff University's Ethics Committee (EC.16.06.14.4547). A total of 901 safety incidents were recorded from a single British rail franchise between 1st June 2010 – 31st December 2016, into the Safety Management Information System (SMIS). The safety incident types identified that fatigue could have been a contributing factor were: Signal Passed at Danger (SPAD), Train Protection & Warning System (TPWS) activation, Automatic Warning System (AWS) slow to cancel, failed to call, and station overrun. From the 901 recorded safety incidents, the scheduling system CrewPlan was used to extract the Fatigue Index (FI) score and Risk Index (RI) score of all train drivers on the day of the safety incident. Only the working rosters of 64.2% (N = 578) (550 men and 28 female) ranging in age from 24 – 65 years old (M = 47.13, SD = 7.30) were accessible for analyses. Analysis from all 578 train drivers who were involved in safety incidents revealed that 99.8% (N = 577) of Fatigue Index (FI) scores fell within or below the identified guideline threshold of 45 as well as 97.9% (N = 566) of Risk Index (RI) scores falling below the 1.6 threshold range. There scores represent good practice within the rail industry. These findings seem to indicate that the current objective indicator, i.e. the FRI calculator used in this study by the British rail franchise was not an effective predictor of train driver's FI scores and RI scores, as safety incidents in which fatigue could have been a contributing factor represented only 0.2% of FI scores and 2.1% of RI scores. Further research is needed to determine whether there are other contributing factors that could provide a better indication as to why there is such a significantly large proportion of train drivers who are involved in safety incidents, in which fatigue could have been a contributing factor have such low FI and RI scores.

***Keywords*—**Fatigue risk index calculator, objective indicator of fatigue, rail industry, safety incident.

Michael Scott Evans is with the School of Psychology, Cardiff University, Cardiff, CF10 3AT UK (phone: +44 2920876599; fax: +44 2920874758; e-mail: evansms3@cardiff.ac.uk).

Andrew Paul Smith is with the School of Psychology, Cardiff University, Cardiff, CF10 3AT UK (phone: +44 2920874757; fax: +44 2920874758; e-mail: smithap@cardiff.ac.uk).

Developing an Objective Indicator of Fatigue: An Alternative Mobile Version of the Psychomotor Vigilance Task (m-PVT)

Michael Scott Evans1[0000-0003-0901-2190], Daniel Harborne2[0000-0002-2376-8341] and Andrew P. Smith1[0000-0001-8805-8028]

[1] School of Psychology, Cardiff University,
63 Park Place, Cardiff, CF10 3AS, Wales, United Kingdom
EvansMS3@cardiff.ac.uk
[2] School of Computer Science and Informatics, Cardiff University,
Cardiff, Wales, UK

Abstract. Approximately 20% of the working population report symptoms of feeling fatigued at work. The aim of the study was to investigate whether an alternative mobile version of the 'gold standard' Psychomotor Vigilance Task (PVT) could be used to provide an objective indicator of fatigue in staff working in applied safety critical settings such as train driving, hospital staffs, emergency services, law enforcements, etc., using different mobile devices. 26 participants mean age 20 years completed a 25-minute reaction time study using an alternative mobile version of the Psychomotor Vigilance Task (m-PVT) that was implemented on either an Apple iPhone 6s Plus or a Samsung Galaxy Tab 4. Participants attended two sessions: a morning and an afternoon session held on two consecutive days counterbalanced. It was found that the iPhone 6s Plus generated both mean speed responses (1/RTs) and mean reaction times (RTs) that were comparable to those observed in the literature while the Galaxy Tab 4 generated significantly lower 1/RTs and slower RTs than those found with the iPhone 6s Plus. Furthermore, it was also found that the iPhone 6s Plus was sensitive enough to detect lower mean speed of responses (1/RTs) and significantly slower mean reaction times (RTs) after 10-minutes on the m-PVT. In contrast, it was also found that the Galaxy Tab 4 generated mean number of lapses that were significant after 5-minutes on the m-PVT. These findings seem to indicate that the m-PVT could be used to provide an objective indicator of fatigue in staff working in applied safety critical settings such as train driving, hospital staffs, emergency services, law enforcements, etc.

Keywords: Psychomotor Vigilance Task (PVT), Mental Workload, Occupational Fatigue, Objective Indicator of Fatigue, Attention.

1 Introduction

In order to be able to meet task demands, there is usually a required amount of operator resources needed, referred to as human mental workload [1]. According to Hart and

Staveland [2], human mental workload can be defined as a ‘cost incurred by a human operator to achieve a particular level of performance’ and evolves from interactions between task demands, circumstances, skills, behaviour, and perceptions.’ Therefore, human mental workload – often referred to as cognitive load – can be intuitively defined as the amount of mental work necessary for a person to complete a task over a given period of time [3, 4]. However, nowadays human mental workload is more generally defined as the measurement of the amount of mental resources involved in a cognitive task [5].

Human mental workload can be measured in real time using a variety of psychological and physiological techniques, which include; subjective psychological self-reported measures e.g., the NASA Task Load Index (NASA-TLX) [2, 6–8] and the NASA-MATB (National Aeronautics and Space Administration Multi-Attribute Task Battery [9] as well as objective physiological measures e.g., heart rate (HR), galvanic skin response (GSR), body temperature, electrocardiogram (ECG), electroencephalogram (EEG), and eye tracking [8, 10–19], and which have been extensively examined in various safety critical environments including; aviation [7, 20], train driving [21], car driving [22–24], and in an operating theater [6] but to name a few.

According to Wickens [25], the greatest value of conducting scientific human mental workload research is to be able to predict the consequences of high mental workload on performance. In other words, to better understand an individual's decision to consciously engage in a safe behaviour or in a potentially dangerous behaviour that could have devastating consequences. As a result, the concept of human mental workload has long been recognised as an important factor in individual performance [26–29]. Xie and Salvendy [29] state that both underload (i.e., low mental workload) and overload (i.e., high mental workload) degrade performance, whereby high and low levels of human mental workload have been shown to lead to operator error [22]. Longo [3] outlines that during low mental workload, individuals are more likely to experience levels of frustration and annoyance when processing information, which could result in an increase in their reaction time (RT). In contract, during high mental workload, individuals could experience confusion, which may result in a decrease in their information processing capacity, which could directly increase the likelihood of errors and mistakes. Therefore, these low and high mental workload information processing stages could have potentially dangerous consequences, especially in safety critical environments. Byrne [30] points out that the main application of mental workload has been to investigate situations where cognitive demand exceeds the acceptable safety tolerance threshold so that workload can be effectively reduced. Therefore, in high risk safety critical environments, the measurement of mental workload is of upmost importance due to its potential implications [31]. However, Xie and Salvendy [29] identified that the effect of fatigue on mental workload is not often considered in human mental workload research. Nevertheless, research carried out by Smith and Smith [32] on conductors/guards and engineers from the rail industry who work in high risk safety critical environments found that workload increased fatigue. However, subjective measures were predominately used in Smith and Smith’s study. As a result, there is a need for an alternative mobile objective indicator of fatigue that can be used in high risk safety critical environments. In a controlled laboratory setting, the human Psychomotor Vigilance Task (PVT) [see 33, 34, for review] has become the widely accepted ‘gold standard’ tool for assessing the impact of fatigue on human cognitive neurobehavioral performance for monitoring temporal dynamic changes in attention [35–38]. The aim of the study was to investigate whether an alternative mobile version of the ‘gold standard’ Psychomotor Vigilance Task (PVT) could be used to provide an objective indicator of fatigue in staff working in applied safety critical settings, such as train driving, hospital staffs, emergency services, law enforcements, etc.

The rest of the paper is organised as follows. Section 2 describes related work on the Psychomotor Vigilance Task (PVT) while also extracting relevant studies to identify the gaps and rationale for the need of an alternative objective indicator of fatigue in staff working in applied safety critical settings. Section 3 outlines the design and empirical methodology of the proposed alternative mobile Psychomotor Vigilance Task (m-PVT). Section 4 presents the empirical results and discussion of the m-PVT. Finally, Section 5 provides a critical conclusion of the proposed alternative m-PVT and suggestions for future work.

2 Related Work

The Psychomotor Vigilance Task (PVT) can be traced back from the early work in simple reaction time (SRT) studies that were carried out by Wilhelm Maximilian Wundt (1832 – 1920) and continued by James McKeen Cattell (1860 – 1944) [39]. It is important to note that the modern PVT has been refined several times over the years [40–42] from its original development by Dinges and Powell [33] and has been shown to be sensitive to sleep deprivation, fatigue, drug use, and age. The PVT has also been widely implemented using a handheld device known as the PVT-192 (Ambulatory Monitoring Inc., Ardsley, New York, USA), as well as being extensively validated by various researchers [40, 43–47].

According to Basner, Mcguire, Goel, Rao and Dinges [48] and Basner et al. [49], the PVT-192 records participants’ sustained attention based on repeated reaction time (RT) trials to visual stimuli that occur at random inter-stimulus intervals (ISI) that are between 2–10 seconds, for a standard 10-minute period. In summary, the PVT-192 device operated by presenting participants with a stimulus that consisted of a four-digit millisecond counter that appears in a light-emitting diode (LED) dot-matrix display. The response consisted of a left or right button press, which depended on the configuration of the PVT-192 setup. The time difference between the stimulus presentation and the response constituted the participant’s reaction time (RT). Each RT value was stored in the device and then uploaded to a personal computer, where the individual RTs are post-processed with the REACT software (Ambulatory Monitoring Inc., Ardsley, New York, USA), or other commercially available software, into summary statistics, such as the mean RT or the mean number of lapses (RTs ≥500 milliseconds) per session [33, 40, 48, 50, 51]. For example, in Roach, Dawson, and Lamond’s study [45], each participant performed either 5 minutes or 10 minutes RT sessions spaced at predetermined intervals (e.g., every 2 hours) for a prolonged duration (e.g., 28 hours), where each session consists of either 50 trials (equivalent to 5 minutes), or 100 trials (equivalent to 10 minutes). However, Khitrov et al. [52] tested the average delay of the PVT-192 and found that the recorded delay was greater than what was stated by the PVT-192 manufacturer. The delay recorded by the researchers was on average 2.4 ms greater when compared to the manufacturer's reported delay of 1 ms. Nevertheless, it is important to highlight that Khitrov et al. [52] did acknowledge the possibility that the difference found could have been due to the non-instantaneous nature of the light detection circuit, or the actual delay associated with the PVT-192, since their experimental design did not permit them to be able to distinguish between these possibilities.

Dinges and Powell [33] have shown that the 10-min PVT is highly reliable. Roach, Dawson and Lamond [45] wanted to investigate whether 90 seconds could also be sufficiently sensitive enough to detect the effects of fatigue in comparison to their earlier research [see 43, for review], where they were able to find significant fatigue-related impairment during the first 5-min of a 10-min PVT. In this study, the researchers compared participants' neurobehavioral performance using the PVT between three different time durations (90 seconds, 5-min, and 10-min) to identify whether a shorter PVT could also be sensitive enough to detect the effects of fatigue. They found that it was only possible to implement a 5-min PVT as a substitute of the 10-min PVT, and not a 90 seconds PVT, thus only further supporting their earlier research [43]. However, it is important to note that analyses of their study were carried out using the mean RT and not the mean speed response (1/RT). Basner and Dinges [43] have identified that the mean RTs should not be the primary measure of alertness, and instead considering using the alternative primary measure of 1/RTs. In a later study, Basner, Mollicone and Dinges [42] aimed to further shorten the 5-min PVT [45] by developing a modified 3-min version of the PVT (PVT-B). They found that this 3-min version could be a useful tool for assessing behavioural alertness in settings where the 'gold standard' 10-min PVT could be more difficult or impractical to implement due to the nature of the study or location. However, further validation is required to determine whether both the 5-min PVT and PVT-B versions could indeed be sensitive enough to detect reduced levels of fatigue. Therefore, this study aimed to investigate a mobile version of the Psychomotor Vigilance Task (m-PVT) that could also be used to provide an objective indicator of fatigue in staff working in applied safety critical settings such as train driving, hospital staffs, emergency services, law enforcements, etc.

3 Design and Methodology

The aim of the study was to investigate whether an alternative mobile version of the 'gold standard' Psychomotor Vigilance Task (PVT) could be used to provide an objective indicator of fatigue in staff working in applied safety critical settings such as train driving, hospital staffs, emergency services, law enforcements, etc. The study received ethics approval from Cardiff University's Ethics Committee (EC.16.02.09.4464R). The study conformed to the seventh amendment of the Declaration of Helsinki 1964 [53] and all participants gave their informed written as well as electronic consent following the explanation of the nature of the study in written form.

3.1 Participants

26 (3 male and 23 female) participants with a mean age of 20 years (SD = 1.66) were recruited as volunteers from Cardiff University via the Experimental Management System (EMS) to take part in the study. The study involved participants attending two sessions, a morning session (i.e., before 11:00) and an afternoon session (i.e., after 17:00), which were held on two consecutive days and counterbalanced, in exchange for £10. The study lasted 60 minutes in total for both sessions.

3.2 Materials / Apparatus

The mobile Psychomotor Vigilance Task (m-PVT) was presented to participants on one of two mobile devices: Apple's iPhone 6s Plus running Apple's iOS version 9.3.1 (Apple Inc.) or Samsung's Galaxy Tab 4 (Samsung Electronics Co. Ltd.) running on Android's operating system (OS) version 4.4.2 KitKat (Alphabet Inc.). The m-PVT ran in the following hardware configurations for the iPhone 6s Plus: system chip (Apple A9 APL1022), processor (Dual-core, 1840 MHz, Twister, 64-bit), graphics processor (PowerVR GT7600), and system memory (2048 MB RAM), and for the Samsung Galaxy Tab 4: system chip (Marvell PXA1088), processor (Quad-core, 1200 MHz, ARM Cortex-A7), graphics processor (Vivante), and system memory (1536 MB RAM). The iPhone 6s Plus had the following hardware configurations: the m-PVT was displayed on either a 5.5-inch (diagonal) 1920 × 1080-pixel native resolution at 401 ppi Retina high definition display (iPhone 6s Plus), or a 7-inch (diagonal) 1280 × 800-pixel (WXGA) native resolution at 216 pixels per inch (ppi) liquid crystal display (LCD) display (Samsung Galaxy Tab 4).

The m-PVT was programmed using the client code HTML, and CSS for the page visualisation and layout. JavaScript was also used to initiate the m-PVT, which was run using the Dolphin Web Browser (MoboTap Inc.) on both an Apple's iPhone 6s Plus and Samsung Galaxy Tab 4 (Dolphin Web Browser versions; Apple app version 9.9.0, and Android app version 11.5.6, respectively). The rationale for selecting the Dolphin Web Browser for this study was that it allowed the full screen feature to be enabled across the two different operating systems (OS), Apple iOS and Android OS platforms for both mobile devices. Other more native mobile internet browsers of each OS platform, such as Safari (Apple) and Chrome (Android) including Firefox, to name a few, did not permit full screen. Qualtrics Surveys (Qualtrics Labs, Inc.) were also used to collect demographic information from participants. These surveys were also implemented on both Apple's iPhone 6s Plus (iOS app version 13.28.06) and Samsung Galaxy Tab 4 (Android app version 1.0.38).

3.3 Statistical Analyses

IBM's Statistical Package for the Social Sciences (SPSS) version 23 for Mac was used to analyse the data. A combination of various statistical procedures were carried out on the data; descriptive analyses, mixed-design analysis of variance (ANOVA) and a two-way analysis of variance (ANOVA) to further explore interactions. The level of $\alpha < .05$ was used for all statistical tests of this experiment.

3.4 Design

The experiment employed a $2 \times 2 \times 6$ mixed-design analysis of variance (ANOVA) with mobile device (Apple's iPhone 6s Plus or Samsung's Galaxy Tab 4) as the between-subjects factor, × time of day (Morning or Afternoon) × time on task (1-minute; 5-minutes; 10-minutes; 15-minutes; 20 minutes; or 25-minutes) as the within-subjects factors. The morning session (i.e., before 11:00) and the afternoon session (i.e., after 17:00) were held on two consecutive days and counterbalanced.

3.5 Procedure

In order to ensure participants were fully aware of the inclusion and exclusion criteria, all participants were contacted using Cardiff University's Experimental Management System (EMS) emailing system 48 hours prior to participation and further reminded 24 hours before the start, in addition to being provided with brief instructions through EMS.

The study was administered using mobile devices. Participants were either assigned to using an iPhone 6s Plus or a Samsung Galaxy Tab 4. To increase validity and standardisation, all instructions were administered to participants in written form for both the morning and the afternoon session. This study consisted of two parts. The first part was the mobile Psychomotor Vigilance Task (m-PVT) reaction time test, which was a modified version of the Dinges and Powell's [33] Psychomotor Vigilance Task. The m-PVT was run on the Dolphin Web Browser mobile application. The second part was the demographic questionnaire that was distributed within Qualtrics Surveys mobile application. In this modified version, the mobile Psychomotor Vigilance Task (m-PVT) (see Figure 1), participants were presented with on-screen instructions and a button at the end that read 'Start'. In each trial, participants were shown a black screen background, and at the centre of the screen they would be presented with a large red fixation circle. The red fixation circle (i.e., inter-stimulus interval) would remain on the screen for a randomised duration that lasted between 2 – 10 seconds, which was then followed by a yellow stimulus counter. As soon as the inter-stimulus interval reached the randomised duration, a yellow stimulus counter appeared counting up in milliseconds from 0 – 5 seconds where it would lapse (i.e., error of omission for 0.5 seconds) and begin the next trial, or until the participant tapped on the screen. Once the participant tapped on the screen, their reaction time (i.e., stimulus) would be displayed for 0.5 seconds. At the end of each trial, a black background would appear on-screen for 0.5 seconds. There were 205 trials in total that lasted approximately 25 minutes. Kribbs and Dinges [54] found that after a maximum of three trials, the practice effect for the PVT was removed. This study conservatively implemented five practice trials to ensure participants were fully aware of the task, which were removed from final analyses. If participants responded prematurely during any trial (i.e., before the timer commenced counting up), the trial would reset. To also ensure participants were made aware of their premature response, the following message in red was displayed on the centre of the screen, 'You clicked too early! This trial will be reset.' A visual illustration of the mobile Psychomotor Vigilance Task (m-PVT) is presented in Figure 2.

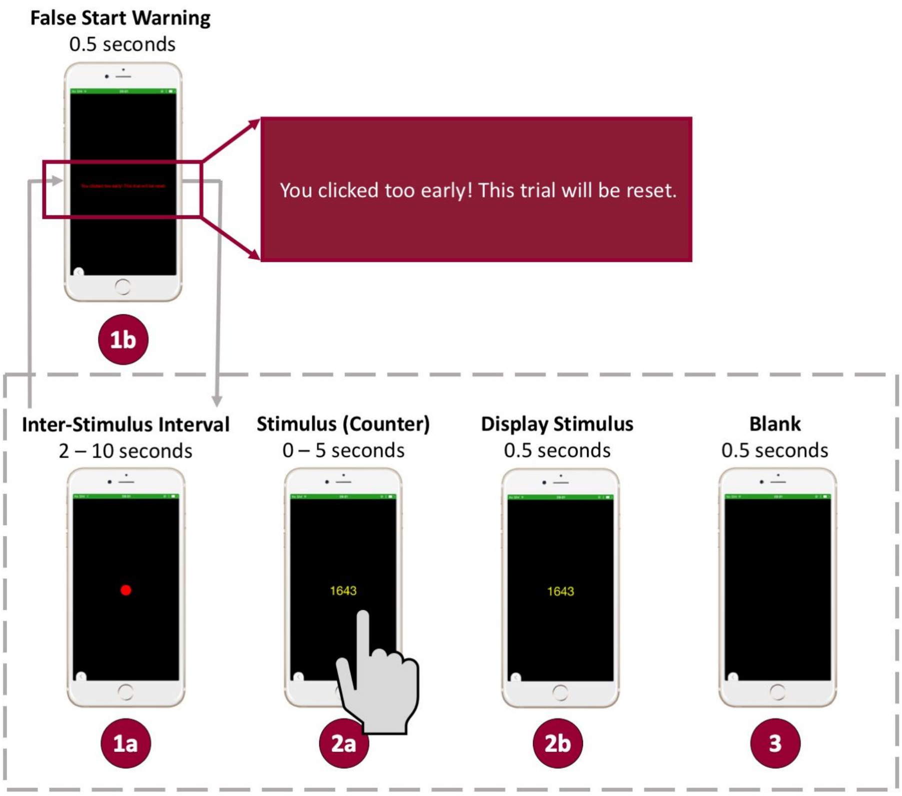


Fig. 1. Mobile Psychomotor Vigilance Task (m-PVT) timeline.

1a. Participants were presented with a large red circle (i.e., inter-stimulus interval), which appeared for a randomised duration between 2 – 10 seconds.

1b. If participants responded prematurely, a false start warning message appeared informing them that they clicked too early and that the trial would be reset.

2a. As soon as the inter-stimulus interval reached the randomised duration, a yellow stimulus counter appeared counting up in milliseconds from 0 – 5 seconds where it would lapse (i.e., error of omission for 0.5 seconds) and begin the next trial, or until the participant had tapped on the screen.

2b. Once the participants had tapped on the screen, their reaction time (i.e., stimulus) would be displayed for 0.5 seconds.

3. At the end of each trial, a black background would appear on-screen for 0.5 seconds.

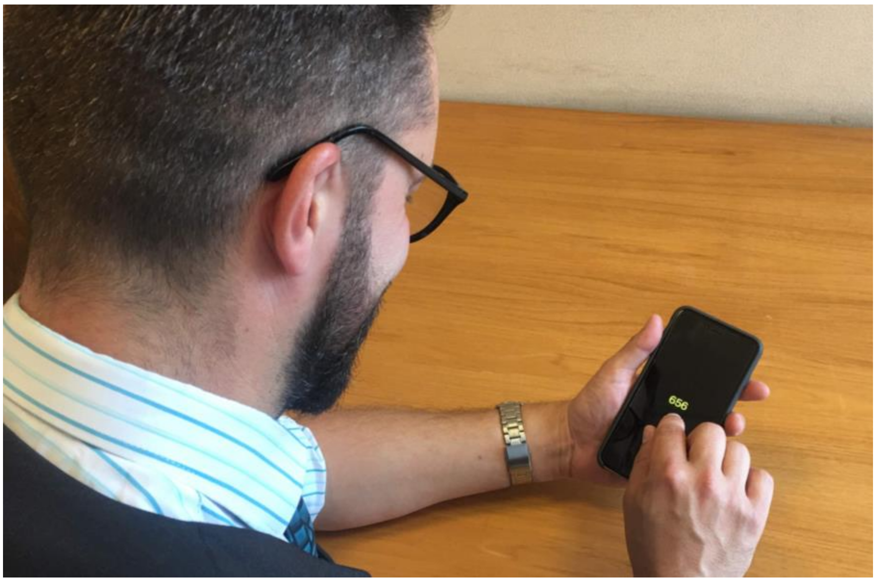


Fig. 2. Visual illustration of the mobile Psychomotor Vigilance Task (m-PVT)

4 Results and Discussion

The aim of the study was to investigate whether an alternative mobile version of the 'gold standard' Psychomotor Vigilance Task (PVT) could be used to provide an objective indicator of fatigue in staff working in applied safety critical settings such as train driving, hospital staffs, emergency services, law enforcerments, etc. IBM's Statistical Package for the Social Sciences (SPSS) version 23 for Mac was used to analyse the data. A total of 10,452 test trials were submitted for data analyses, with all 260 practice trials (i.e., 5 practice trials per session) excluded from final analyses. It is important to note that all mobile devices running the online mobile version of the Psychomotor Vigilance Task (m-PVT) were administered through the Dolphin internet browser and were connected using Cardiff University's Eduroam Wi-Fi roaming service. Therefore, on rare occasions when the Wi-Fi connectivity dropped, the participant's trial was lost and thus not recorded. As a result, a total of 1.95% ($n = 208$) test trials of all 10,660 trials (i.e., 260 practice and 10,400 test) were lost and not recorded. Based on Basner and Dinges [40] recommendations, all 10,452 test trials with reaction time (RTs) < 100 ms (i.e., false start), which accounted for .05% ($n = 5$) and RTs $\geq$ 500 ms (i.e., number of lapses), which accounted for 31.84% ($n = 3,328$), were considered for exclusion from the final mean speed response (1/RT) and mean reaction time (RT) analyses. All 31.84% ($n = 3,328$) of RTs $\geq$ 500 ms (i.e., number of lapses) were analysed separately.

4.1 Mean Speed Response (1/RT) and Reaction Time (RT)

Figure 3 presents the illustrated mean speed responses (1/RTs) across the different conditions while Figure 4 presents the illustrated mean reaction times (RTs) across the different conditions. Both the 1/RTs and RTs were submitted to a 2 × 2 × 6 mixed-design analysis of variance (ANOVA) with 2 × mobile devices (iPhone 6s Plus or Samsung Galaxy Tab 4) as the between-subjects factor, and × 2 time of day (Morning, or Afternoon) × 6 time on task (1-minute; 5-minutes, 10-minutes, 15-minutes, 20 minutes, or 25-minutes) as the within-subjects factors. Both the 1/RTs and RTs were significant when comparing the main effect of the two groups using different mobile devices, $F(1, 24)$, 87.21, $p < .001$, $\eta_p^2 = .78$, indicating a large effect size [55, 56] and $F(1, 24)$, 131.85, $p < .001$, $\eta_p^2 = .85$, also indicating a large effect size [55, 56], respectively. In addition, there was a significant main effect of time on task for both the 1/RTs and RTs, Wilks' Lambda = .22, $F(5, 20)$, 14.08, $p < .001$, $\eta_p^2 = .78$, indicating a large effect size [55, 56] and Wilks' Lambda = .24, $F(5, 20)$, 12.66, $p < .001$, $\eta_p^2 = .76$, indicating a large effect size [55, 56], respectively. Furthermore, there was also a significant interaction between mobile devices × time on task for both the 1/RTs and RTs, Wilks' Lambda = .34, $F(5, 20)$, 7.95, $p < .001$, $\eta_p^2 = .67$, indicating a large effect size [55, 56] and Wilks' Lambda = .43, $F(5, 20)$, 5.23, $p = .003$, $\eta_p^2 = .57$, indicating a moderate effect size [55, 56], respectively. The other main effect (time of day) and interactions (two-way interaction, time of day × time on task; and three-way interaction, mobile devices × time of day × time on task) for both 1/RTs and RTs were not significant.

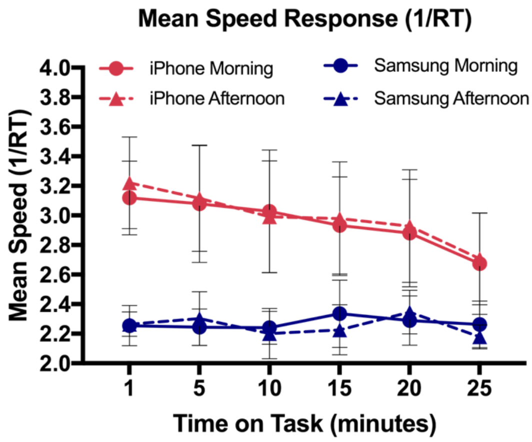


Fig. 3. Mean speed responses (1/RTs) across the different conditions (i.e., morning and afternoon) for both the iPhone 6s Plus and the Samsung Galaxy Tab 4 of the mobile Psychomotor Vigilance Task (m-PVT). *Note:* Mean 1/RTs for both the iPhone 6s Plus and the Samsung Galaxy Tab 4 are presented in bins of 5 minutes as well as the first minute. Error bars represents standard deviation.

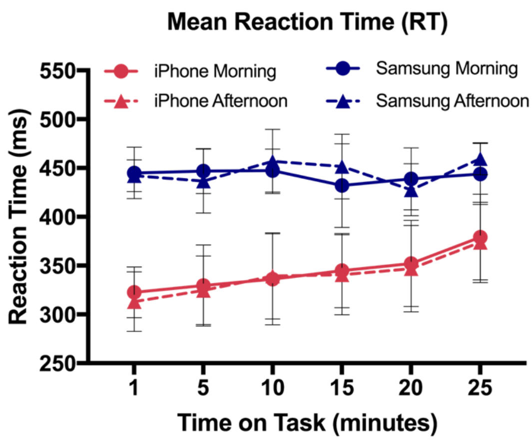


Fig. 4. Mean reaction times (RTs) across the different conditions (i.e., morning and afternoon) for both the iPhone 6s Plus and the Samsung Galaxy Tab 4 of the mobile Psychomotor Vigilance Task (m-PVT). *Note:* Mean RTs for both the iPhone 6s Plus and the Samsung Galaxy Tab 4 are presented in bins of 5 minutes as well as the first minute. Error bars represents standard deviation.

The main effect of the two groups using different mobile devices was followed by post-hoc tests with Bonferroni correction for multiple comparisons. Post-hoc tests showed that participants' mean speed responses (1/RTs) were significantly greater with the iPhone 6s Plus mobile device (M = 2.97, SE = .05) than the Samsung Galaxy Tab 4 mobile device (M = 2.26, SE = .05, p < .001). In addition, post-hoc tests also showed that participants' reaction times (RTs) were significantly faster with the iPhone 6s Plus mobile device (M = 341.92ms, SE = 6.29ms) than the Samsung Galaxy Tab 4 mobile device (M = 444.02ms, SE = 6.29ms, p < .001). These findings seem to indicate that the iPhone 6s Plus generated significantly greater mean speed responses (1/RTs) and significantly faster mean reaction times (RTs) than the Samsung Galaxy Tab 4, with a mean RT difference of 102ms between the iPhone 6s Plus and the Samsung Galaxy Tab 4. Therefore, under these circumstances, the interaction between mobile devices × time on task was explored separately with a two-way repeated analysis of variance (ANOVA).

3.1.1: iPhone 6s Plus Mean Speed Response (1/RT) and Reaction Time (RT)

Figure 5 and Figure 6 present the illustrated mean speed of responses (1/RTs) and mean reaction times (RTs) for the iPhone 6s Plus mobile Psychomotor Vigilance Task (m-PVT) across the different conditions. Both the 1/RTs and RTs were submitted to a 2 × 6 two-way repeated analysis of variance (ANOVA) comparing 2 × time of day (Morning, or Afternoon) × 6 time on task (1-minute; 5-minutes, 10-minutes, 15-minutes, 20 minutes, or 25-minutes). Only the main effect of time on task was significant for both the 1/RTs and RTs, Wilks' Lambda = .12, $F(5, 8)$, 12.02, p = .001, η_p^2 = .88, indicating a large effect size [55, 56] and Wilks' Lambda = .12, $F(5, 8)$, 11.93, $p = .002$, $\eta_p^2 = .88$, indicating a large effect size [55, 56], respectively. The other main effect (time of day) and interactions (two-way interaction, time of day × time on task) for both 1/RTs and RTs were not significant.

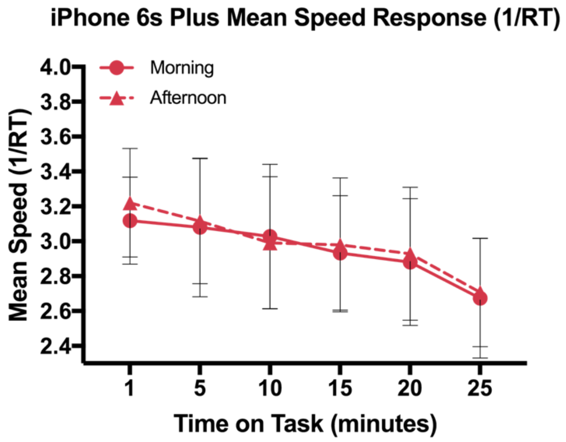


Fig. 5. Mean speed responses (1/RTs) of both the morning session and afternoon session for the iPhone 6s Plus mobile Psychomotor Vigilance Task (m-PVT). *Note:* Mean 1/RTs of both the morning session and afternoon session for the iPhone 6s Plus are presented in bins of 5 minutes as well as the first minute. Error bars represents standard deviation.

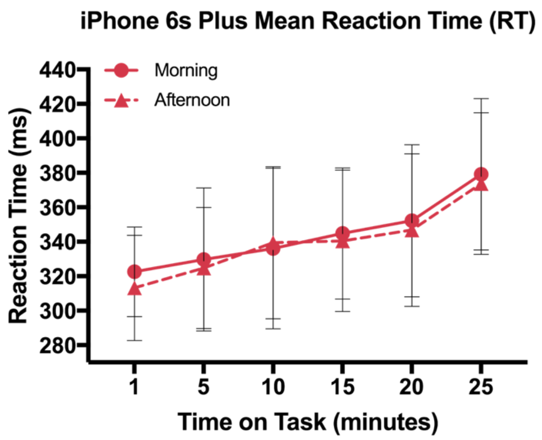


Fig. 6. Mean reaction times (RTs) of both the morning session and afternoon session for the iPhone 6s Plus mobile Psychomotor Vigilance Task (m-PVT). *Note:* Mean RTs of both the morning session and afternoon session are presented in bins of 5 minutes as well as the first minute. Error bars represents standard deviation.

The main effect of time on task was further explored using Fisher's Least Significant Difference (LSD) post-hoc multiple pairwise comparison, which according to Rovai, Baker and Ponton [57] is used when sample sizes are small. As can be seen from Figure

7, participants who were assigned to the iPhone mobile device group had significantly greater mean speed responses (1/RTs) between the first minute on the m-PVT (M = 3.17, SE = .07) and 15-minutes on the m-PVT (M = 2.96, SE = .09, p = .005). In addition, participants had significantly greater 1/RTs between the first minute (M = 3.17, SE = .07) and 20-minutes (M = 2.90, SE = .10, p = .005). Furthermore, participants had significantly greater 1/RTs between the first minute (M = 3.17, SE = .07) and 25-minutes (M = 2.69, SE =.07, p < .001). Fisher's LSD post-hoc multiple pairwise comparison also showed potential differences between the first minute on the m-PVT (M = 3.17, SE = .07) and 10-minutes on the m-PVT (M = 3.01, SE = .10, p = .051). However, this was not statistically significant with this study size. As can be seen from Figure 8, participants had significantly faster mean reaction times (RTs) between the first minute on the m-PVT (M = 317.89ms, SE = 7.09ms) and 10-minutes on the m-PVT (M = 337.75ms, SE = 10.27ms, p = .032). In addition, participants had significantly faster RTs between the first minute (M = 317.89ms, SE = 7.09ms) and 15-minutes (M = 342.70ms, SE = 10.22ms, p = .003). Furthermore, participants had significantly faster RTs between the first minute (M = 317.89ms, SE = 7.09ms) and 20-minutes (M = 349.52ms, SE = 11.42ms, p = .005). Moreover, participants had significantly faster RTs between the first minute (M = 317.89ms, SE = 7.09ms) and 25-minutes (M = 376.47ms, SE = 9.20ms, p < .001).

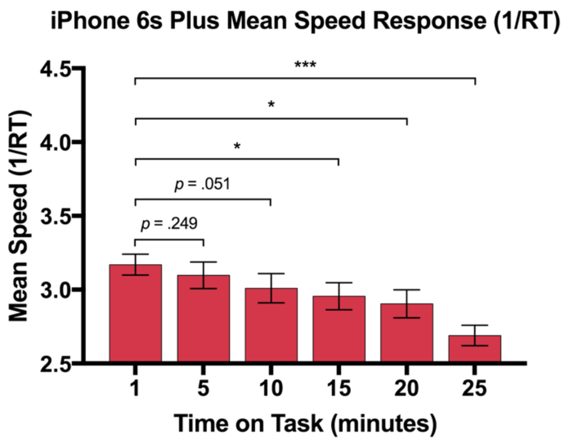


Fig. 7. *p < .05; **p < .005; ***p < .001. *Note:* Mean speed responses (1/RTs) for the iPhone 6s Plus are presented in bins of 5 minutes as well as the first minute. Error bars represents standard errors.

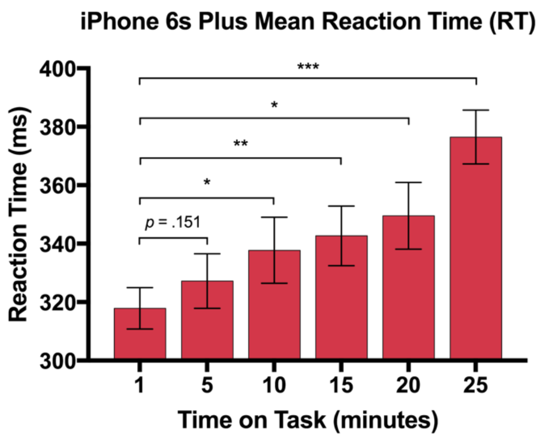


Fig. 8. *p < .05; ***p* < .005; ****p* < .001. *Note:* Mean reaction times (RTs) for the iPhone 6s Plus are presented in bins of 5 minutes as well as the first minute. Error bars represents standard errors.

3.1.2: Samsung Galaxy Tab 4 Mean Speed Response (1/RT) and Reaction Time (RT)

Figure 9 and Figure 10 present the illustrated mean speed responses (1/RTs) and mean reaction times (RTs) for Samsung Galaxy Tab 4 mobile Psychomotor Vigilance Task (m-PVT) across the different conditions. Both the 1/RTs and RTs were submitted to a 2 × 6 two-way repeated analysis of variance (ANOVA) comparing 2 × time of day (Morning, or Afternoon) × 6 time on task (1-minute; 5-minutes, 10-minutes, 15-minutes, 20 minutes, or 25-minutes). For both the 1/RTs and RTs, there was no significant main effect of time of day; Wilks' Lambda = .96, $F(1, 12)$, .530, $p = .481$, $\eta_p^2 = .04$ and Wilks' Lambda = .95, $F(1, 12)$, .579, $p = .461$, $\eta_p^2 = .05$, respectively. In addition, for both the 1/RTs and RTs, there was also no significant main effect of time on task; Wilks' Lambda = .31, $F(5, 8)$, 3.56, $p = .054$, $\eta_p^2 = .69$ and Wilks' Lambda = .31, $F(5, 8)$, 3.53, $p = .056$, $\eta_p^2 = .69$, respectively. Moreover, for both the 1/RTs and RTs, there was also no significant interaction between time of day × time of task; Wilks' Lambda = .61, $F(5, 8)$, 1.05, $p = .454$, $\eta_p^2 = .40$ and Wilks' Lambda = .63, $F(5, 8)$, .954, $p = .497$, $\eta_p^2 = .37$, respectively.

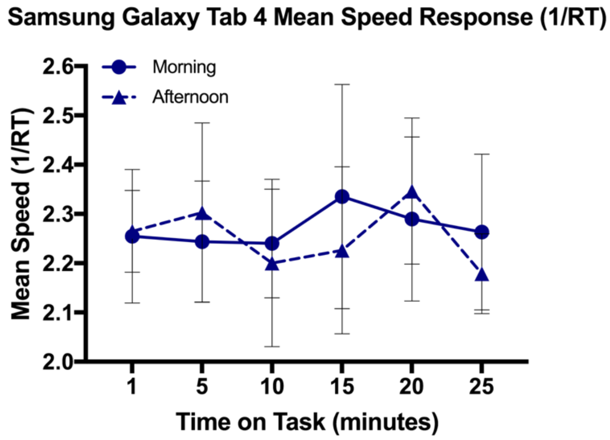


Fig. 9. Mean speed responses (1/RTs) of both the morning session and afternoon session of the Samsung Galaxy Tab 4 mobile Psychomotor Vigilance Task (m-PVT). *Note:* Mean 1/RTs of both the morning session and afternoon session for the Samsung Galaxy Tab 4 are presented in bins of 5 minutes as well as the first minute. Error bars represents standard deviation.

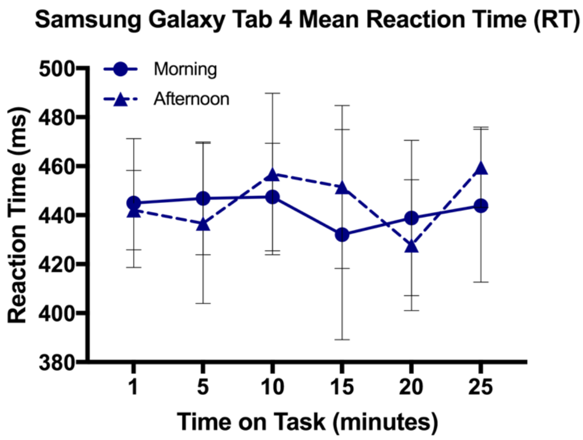


Fig. 10. Mean reaction times (RTs) of both the morning session and afternoon session of the Samsung Galaxy Tab 4 mobile Psychomotor Vigilance Task (m-PVT). *Note:* Mean RTs of both the morning session and afternoon session for the Samsung Galaxy Tab 4 are presented in bins of 5 minutes as well as the first minute. Error bars represents standard deviation.

4.2 Mean Number of Lapses

From all test trials, a total of 31.84% (n = 3,328) RTs ≥ 500 ms were submitted for data analyses. Figure 11 presents the illustrated mean number of lapses across the different conditions. The mean number of lapses were submitted to a 2 × 2 × 6 mixed-design analysis of variance (ANOVA) with 2 × mobile devices (iPhone 6s Plus or Samsung

Galaxy Tab 4) as the between-subjects factor, and × 2 time of day (Morning, or Afternoon) × 6 time on task (1-minute; 5-minutes, 10-minutes, 15-minutes, 20 minutes, or 25-minutes) as the within-subjects factors. There was a significant main effect of the two groups using different mobile devices, $F(1, 24)$, 131.81, $p < .001$, $\eta_p^2 = .85$, indicating a large effect size [55, 56]. In addition, there was a significant main effect of time on task, Wilks' Lambda = .28, $F(5, 20)$, 10.27, $p < .001$, $\eta_p^2 = .72$, indicating a large effect size [55, 56]. Furthermore, there was also a significant interaction between mobile devices × time on task, Wilks' Lambda = .31, $F(5, 20)$, 9.10, $p < .001$, $\eta_p^2 = .70$, indicating a large effect size [55, 56]. The other main effect (time of day, $p = .620$) and interactions (two-way interaction, time of day × time on task, $p = .395$; and three-way interaction, mobile devices × time of day × time on task, $p = .151$) for the mean number of lapses (i.e., RTs ≥ 500 ms) were not significant.

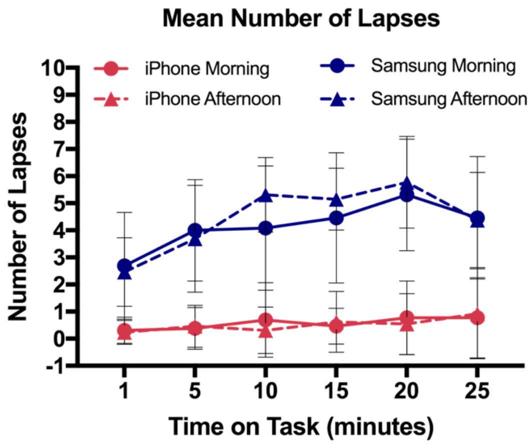


Fig. 11. Mean number of lapses across the different conditions (i.e., morning and afternoon) for both the iPhone 6s Plus and the Samsung Galaxy Tab 4 of the mobile Psychomotor Vigilance Task (m-PVT). *Note:* Mean number of lapses for both the iPhone 6s Plus and the Samsung Galaxy Tab 4 are presented in bins of 5 minutes as well as the first minute. Error bars represents standard deviation.

The main effect of the two groups using different mobile devices was followed by post-hoc tests with Bonferroni correction for multiple comparisons. Post-hoc tests showed that participants' mean number of lapses were significantly lower for the iPhone 6s Plus mobile device ($M = .54$, $SE = .23$) than the Samsung Galaxy Tab 4 mobile device ($M = 4.31$, $SE = .23$, $p < .001$). These findings seem to indicate that participants assigned to the iPhone 6s Plus recorded significantly less mean number of lapses than the Samsung Galaxy Tab 4. These findings are not too surprising as it was previously found that both the mean speed responses (1/RTs) and mean reaction times (RTs) for the iPhone 6s Plus generated significantly greater 1/RTs and faster RTs than the Samsung Galaxy Tab 4. There was a statistically difference of 102ms, which would indicate at least for the Samsung Galaxy Tab 4 that there would be significantly more test trials with RTs ≥ 500 ms (i.e., number of lapses). As a result, from all 31.84% (n = 3,328) of test trials with RTs ≥ 500 ms, the Samsung Galaxy Tab 4 group represented 90.32% (n = 3,006) and the iPhone 6s Plus group represented 9.68% (n = 322). Therefore, also under these circumstances, the interaction between mobile devices × time on task was explored separately with a two-way repeated analysis of variance (ANOVA).

3.1.3: iPhone 6s Plus Mean Number of Lapses

Figure 12 presents the illustrated mean number of lapses for the iPhone 6s Plus mobile Psychomotor Vigilance Task (m-PVT) across the different conditions. The mean number of lapses were submitted to a 2 × 6 two-way repeated analysis of variance (ANOVA) comparing 2 × time of day (Morning, or Afternoon) × 6 time on task (1-minute; 5-minutes, 10-minutes, 15-minutes, 20 minutes, or 25-minutes). There was no significant main effect of time of day; Wilks' Lambda = .997, $F(1, 12)$, .04, $p = .846$, $\eta_p^2 = .00$. In addition, there was also no significant main effect of time on task; Wilks' Lambda = .75, $F(5, 8)$, .54, $p = .744$, $\eta_p^2 = .25$. Moreover, there was also no significant interaction between time of task × time of day; Wilks' Lambda = .36, $F(5, 8)$, 2.84, $p = .092$, $\eta_p^2 = .64$.

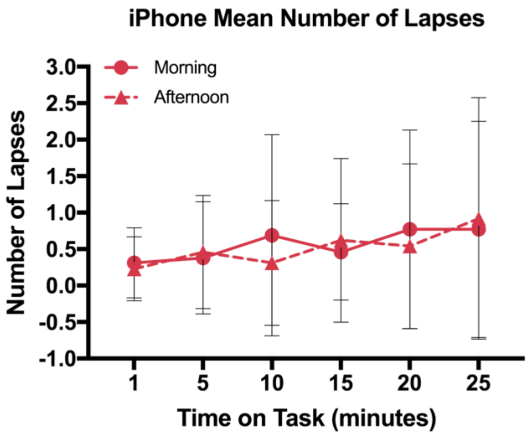


Fig. 12. Mean number of lapses for both the morning session and afternoon session for the iPhone 6s Plus of the mobile Psychomotor Vigilance Task (m-PVT). *Note:* Mean number of lapses for the iPhone 6s Plus are presented in bins of 5 minutes as well as the first minute. Error bars represents standard deviation.

3.1.4: Samsung Galaxy Tab 4 Mean Number of Lapses

Figure 13 presents the illustrated mean number of lapses for the Samsung Galaxy Tab 4 mobile Psychomotor Vigilance Task (m-PVT) across the different conditions. The mean number of lapses were submitted to a 2 × 6 two-way repeated analysis of variance (ANOVA) comparing 2 × time of day (Morning, or Afternoon) × 6 time on task (1-minute; 5-minutes, 10-minutes, 15-minutes, 20 minutes, or 25-minutes). Only the main effect of time on task was significant for the mean number of lapses, Wilks' Lambda = .14, *F*(5, 8), 9.80, *p* = .003, η_p^2 = .86, indicating a large effect size [55, 56]. The other main effect (time of day, *p* = .486) and two-way interaction (time of day × time on task, *p* = .227) for the mean number of lapses (i.e., RTs ≥ 500 ms) were not significant.

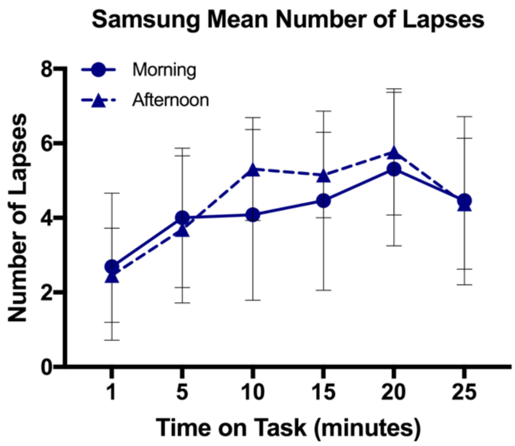


Fig. 13. Mean number of lapses for both the morning session and afternoon session for the Samsung Galaxy Tab 4 of the mobile Psychomotor Vigilance Task (m-PVT). *Note:* Mean number of lapses for the Samsung Galaxy Tab 4 are presented in bins of 5 minutes as well as the first minute. Error bars represents standard deviation.

The main effect of time on task was further explored using Fisher's Least Significant Difference (LSD) post-hoc multiple pairwise comparison, which according to Rovai, Baker and Ponton [57] is used when sample sizes are small. As can be seen from Figure 14, participants who were assigned to the Samsung Galaxy Tab 4 mobile device group had significantly less mean number of lapses between the first minute on the m-PVT (*M* = 2.58, *SE* = .35) and 5-minutes on the m-PVT (*M* = 3.85, *SE* = .37, *p* = .001). In addition, participants also had significantly less mean number of lapses between the first minute on the m-PVT (*M* = 2.58, *SE* = .35) and 10-minutes on the m-PVT (*M* = 4.69, *SE* = .40, *p* < .001). Furthermore, participants also had significantly less mean number of lapses between the first minute on the m-PVT (*M* = 2.58, *SE* = .35) and 15-minutes on the m-PVT (*M* = 4.81, *SE* = .40, *p* = .001). Moreover, participants also had a significantly lower mean number of lapses between the first minute on the m-PVT (*M* = 2.58, *SE* = .35) and 20-minutes on the m-PVT (*M* = 5.54, *SE* = .38, *p* < .001). Finally, participants also had a significantly lower mean number of lapses between the first minute on the m-PVT (*M* = 2.58, *SE* = .35) and 25-minutes on the m-PVT (*M* = 4.42, *SE* = .46, *p* = .008). These findings seem to indicate that mean number of lapses for mobile devices, that generate on average significantly slower thresholds, due to perhaps hardware configurations than what is typically found in the Psychomotor Vigilance Task (PVT) literature, may not be an accurate representation and comparison from analyses of both the mean speed responses (1/RT) and mean reaction times (RTs). Instead, the analyses of the mean number of lapses may yield far better research insights.

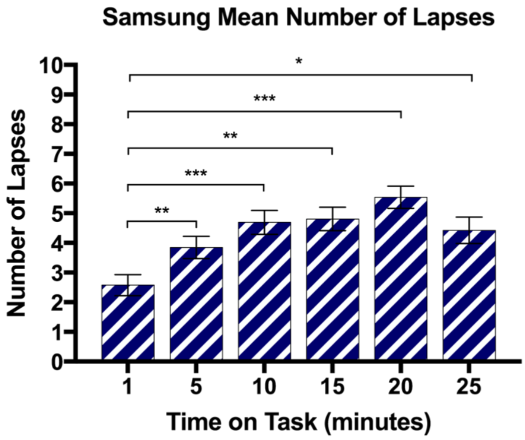


Fig. 14. *p < .05; ***p* < .005; ****p* < .001. *Note:* Mean number of lapses for the Samsung Galaxy Tab 4 are presented in bins of 5 minutes as well as the first minute. Error bars represents standard errors.

5 Conclusion

The study aimed to investigate whether an alternative online mobile version of the 'gold standard' Psychomotor Vigilance Task (PVT) could be used to provide an objective indicator of fatigue in staff in applied safety critical settings such as train driving, hospital staffs, emergency services, law enforcements, etc. It was found that there was a large significant difference in reaction times (RTs) between the two mobile devices (i.e., Samsung vs. Apple's iPhone). Apple's iPhone 6s Plus generated RTs that were comparable to those found in the literature [33, 34, 40, 42–46, 52]. However, the RTs of the Samsung mobile device were significantly slower than those found in the literature. Findings from this study also support previous research that have identified that an increase in fatigue results in impaired alertness [58, 59], whereby sustained attention, as measured by reaction time, significantly reduces after 10-minutes of continuous performance using the Psychomotor Vigilance Task (PVT). These findings from this alternative online mobile version of the Psychomotor Vigilance Task (m-PVT) are consistent with previous work, which suggested that sustained attention drops with prolonged duration of the task [60. 61].

This study seems to suggest that an alternative online mobile version of the 'gold standard' 10-min PVT (i.e., m-PVT) could be used to provide an objective indicator of fatigue after 10 minutes on the m-PVT in staff working in applied safety critical settings such as train driving, hospital staffs, emergency services, law enforcments, etc. However, caution is required when considering implementing an alternative online mobile version (m-PVT) that is running on an internet browser, as only the iPhone 6s Plus was able to generate reaction times that were comparable with the literature. In contrast, there were significantly fewer lapses for the iPhone 6s Plus (n = 322) than the Samsung Galaxy Tab 4 (n = 3,006), which was not surprising when considering that both mean speed responses (1/RTs) and reaction times (RTs) were significantly higher and faster respectively, for the iPhone 6s Plus than for the Samsung Galaxy Tab 4. As a result, perhaps analyses of both the mean speed responses (1/RTs) and mean reaction times (RTs) may not always generate an accurate data representation for analyses based on the hardware differences in mobile manufactures as well as configurations and specifications. Therefore, perhaps using the number of lapses (i.e., RTs ≥ 500 ms) may yield richer data for analyses on these circumstances. As a result, this study recommends that pilot studies should be carried out to firstly explore and determine whether the selected mobile device generates RTs that are better suited for either mean RTs and mean 1/RTs, or mean number of lapses analyses. However, there are several factors that could also account for the difference in the mean 1/RTs and mean RTs between the two mobile devices.

Firstly, regarding software, both the Apple's iPhone 6s Plus and Samsung Galaxy Tab 4 run on different operating systems (OS). Apple's iPhone 6s Plus run their own native iOS version 9.3.1, while the Samsung Galaxy Tab 4 run on Alphabet's Android KitKat version 4.4.2. Furthermore, even though the same internet browser (Dolphin Web Browser) was used across both mobile devices, the version numbers were different. This may indicate that one may have had more improvement and stability updates than the other (Dolphin Web Browser; Apple\s native iOS app version 9.9.0 vs. Android app OS version 11.5.6). Alternatively, the browser may have been developed for one platform and then expanded to also run on the other platform.

5.1 Future Work

Further research is now needed to determine whether the m-PVT can be used to provide an objective indicator of fatigue in staff working in applied safety critical settings such as train driving, hospital staffs, emergency services, law enforcement, etc. Use of an iPhone 6s Plus is recommended, and further studies with larger samples are required to confirm the length of the task.

6 Acknowledgment

This work was supported by the Economic and Social Research Council [ES/J500197/1] and Arriva Trains Wales (ATW).

7 Contributors

MSE designed the experiment, carried out data collection, statistical analysis and writing of the paper. DH programmed the Psychomotor Vigilance Task (m-PVT) under the directions of MSE using the client code HTML, and CSS for the page visualisation and layout. DH also used the JavaScript to initiate the m-PVT. MSE wrote the paper with input from APS. All authors have seen and approved the final version of the paper for publication.

8 References

1. Eggemeier, F.T., Wilson, G.F., Kramer, A.F., Damos, D.L.: General considerations concerning workload assessment in multi-task environments. In: Damos, D. L. (eds.) Multiple Task Performance. pp. 207–216. Taylor & Francis, London 1991.
2. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (task load index): Results of empirical and theoretical research. Advances in Psychology. **52**, 139–183 1988. doi: 10.1016/S0166-4115(08)62386-9
3. Longo, L.: Designing medical interactive systems via assessment of human mental workload. In: International Symposium on Computer-Based Medical Systems. pp. 364–365 2015. dio: 10.1109/CBMS.2015.67
4. Longo, L.: Subjective Usability, Mental Workload Assessments and Their Impact on Objective Human Performance. In: IFIP Conference on Human-Computer Interaction. pp. 202–223. 2017. dio: 10.1007/978-3-319-67684-5_13
5. Zammouri, A., Moussa, A.A., Mebrouk, Y.: Brain-computer interface for workload estimation: Assessment of mental efforts in learning processes. Expert Systems with Applications. **112**, 138–147 2018. doi: 10.1016/j.eswa.2018.06.027
6. Byrne, A.J., Oliver, M., Bodger, O., Barnett, W.A., Williams, D., Jones, H., Murphy, A.: Novel method of measuring the mental workload of anaesthetists during clinical practice. British Journal of Anaesthesia. **105**, 767–771 2010. doi: 10.1093/bja/aeq240
7. Orlandi, L., Brooks, B.: Measuring mental workload and physiological reactions in marine pilots: Building bridges towards redlines of performance. Applied Ergonomics. **69**, 74–92 2018. doi: 10.1016/j.apergo.2018.01.005
8. Shakouri, M., Ikuma, L.H., Aghazadeh, F., Nahmens, I.: Analysis of the sensitivity of heart rate variability and subjective workload measures in a driving simulator: The case of highway work zones. International Journal of Industrial Ergonomics. **66**, 136–145 2018. doi: 10.1016/j.ergon.2018.02.015
9. Comstock, J.R.J., Arnegard, R.J.: The Multi-Attribute Test Battery for Human Operator Workload and Strategic Behaviour. National Aeronautics and Space Administration, Hampton 1992.
10. Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., Olmstead, R.E., Tremoulet, P.D., Craven, P.L.: EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. Aviation Space and Environmental Medicine. **78**, B231–B244 2007.
11. Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., Babiloni, F.: Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neuroscience and Biobehavioral Reviews. **44**, 58–75 2014. doi: 10.1016/j.neubiorev.2012.10.003
12. Heine, T., Lenis, G., Reichensperger, P., Beran, T., Doessel, O., Deml, B.: Electrocardiographic features for the measurement of drivers' mental workload. Applied Ergonomics. **61**, 31–43. 2017. doi: 10.1016/j.apergo.2016.12.015
13. Hogervorst, M.A., Brouwer, A.M., Van Erp, J.B.E.: Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. Frontiers in Neuroscience. **8**, 322 2014. doi: 10.3389/fnins.2014.00322
14. Hsu, B.W., Wang, M.J.J., Chen, C.Y., Chen, F.: Effective indices for monitoring mental workload while performing multiple tasks. Perceptual and Motor Skills. **121**, 94–117 2015. doi: 10.2466/22.PMS.121c12x5
15. Jimenez-Molina, A., Retamal, C., Lira, H.: Using psychophysiological sensors to assess mental workload during web browsing. Sensors. **18**, 458 2018. doi: 10.3390/s18020458

16. Shaw, E.P., Rietschel, J.C., Hendershot, B.D., Pruziner, A.L., Miller, M.W., Hatfield, B.D., Gentili, R.J.: Measurement of attentional reserve and mental effort for cognitive workload assessment under various task demands during dual-task walking. Biological Psychology. **134**, 39–51 2018. doi: 10.1016/j.biopsycho.2018.01.009
17. So, W.K.Y., Wong, S.W.H., Mak, J.N., Chan, R.H.M.: An evaluation of mental workload with frontal EEG. PLOS ONE. **12**, e0174949 2017. doi: 10.1371/journal.pone.0174949
18. Vergara, R.C., Moenne-Loccoz, C., Maldonado, P.E.: Cold-Blooded attention: Finger temperature predicts attentional performance. Frontiers in Human Neuroscience. **11**, 454 2017. doi: 10.3389/fnhum.2017.00454
19. Widyanti, A., Hanna., Muslim, K., Sutalaksana, I.Z.: The sensitivity of Galvanic Skin Response for assessing mental workload in Indonesia. Work: A Journal of Prevention Assessment & Rehabilitation. **56**, 111–117 2017. doi: 10.3233/WOR-162479
20. Blanco, J.A., Johnson, M.K., Jaquess, K.J., Oh, H., Lo, L.C., Gentili, R.J., Hatfield, B.D.: Quantifying Cognitive Workload in Simulated Flight Using Passive, Dry EEG Measurements. IEEE Transactions on Cognitive and Developmental Systems. **10**, 373–383 2018. doi: 10.1109/TCDS.2016.2628702
21. Myrtek, M., Deutschmannjanicke, E., Strohmaier, H., Zimmermann, W., Lawerenz, S., Brugner, G., Muller, W.: Physical, mental, emotional, and subjective workload components in train drivers. Ergonomics. **37**, 1195–1203 1994. doi: 10.1080/00140139408964897
22. Brookhuis, K.A., De Waard, D.: Assessment of drivers' workload: Performance and subjective and physiological indexes. In: Hancock, P.A., Desmond, P.A. (eds.) Stress, Workload and Fatigue. pp. 321–333. Lawrence Erlbaum Associates, New Jersey 2001.
23. Foy, H.J., Chapman, P.: Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation. Applied Ergonomic. **73**, 90–99 2018. doi: 10.1016/j.apergo.2018.06.006
24. Paxion, J., Galy, E., Berthelon, C.: (2014). Mental workload and driving. Frontiers in Psychology. **5**, 1344 2014. doi: 10.3389/fpsyg.2014.01344
25. Wickens, C.D.: Mental workload: assessment, prediction and consequences. In International Symposium on Human Mental Workload: Models and Applications. pp. 18–29. Springer, Cham 2017. dio: 10.1007/978-3-319-61061-0_2
26. Gopher, D., Donchin, E.: Workload: An examination of the concept. In: Boff, K.R., Kaufman, L., Thomas, J.P. (eds.) Handbook of perception and human performance, Vol. 2. Cognitive processes and performance. pp. 1–49. John Wiley & Sons, Oxford, England 1986.
27. Hancock, P.A., Meshkati, N.: Human Mental Workload. Elsevier Science, Amsterdam 1988.
28. Moray, N.: Mental Workload: Its Theory and Measurement. Plenum, New York 1979.
29. Xie, B., Salvendy, G.: Review and reappraisal of modelling and predicting mental workload in single- and multi-task environments. Work and Stress. **14**, 74–99 2000. doi: 10.1080/026783700417249
30. Byrne, A.: Mental Workload as an Outcome in Medical Education. In International Symposium on Human Mental Workload: Models and Applications. pp. 187–197. Springer, Cham. 2017. doi: doi.org/10.1007/978-3-319-61061-0_12
31. Gaba, D.M., & Lee, T.: (1990). Measuring the workload of the anesthesiologist. Anesthesia and Analgesia. **71**, 354–361 1990.
32. Smith, A. P., Smith, H.: Workload, fatigue and performance in the rail industry. In International Symposium on Human Mental Workload: Models and Applications. pp. 251–263. 2017. Springer, Cham. doi: 10.1007/978-3-319-61061-0_17
33. Dinges, D.F., Powell, J. W.: Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. Behavior Research Methods, Instruments, & Computers. **17**, 652–655 (1985). doi: 10.3758/BF03200977

34. Dinges, D.F., Orne, M.T., Whitehouse, W.G., Orne, E.C.: Temporal placement of a nap for alertness: contributions of circadian phase and prior wakefulness. Sleep. **10**, 313–329 (1987).
35. Belenky, G., Wesensten, N.J., Thorne, D.R., Thomas, M.L., Sing, H.C., Redmond, D.P., Russo, M.B., Balkin, T.J.: Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study. Journal of Sleep Research. **12**, 1–12 (2003). doi: 10.1046/j.1365-2869.2003.00337.x
36. Dinges, D.F., Pack, F., Williams, K., Gillen, K.A., Powell, J.W., Ott, G.E., Aptowicz, C., Pack, A.I.: Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. Sleep. **20**, 267–777 (1997).
37. Jewett, M.E., Dijk, D.J., Kronauer, R.E., Dinges, D.F.: Dose-response relationship between sleep duration and human psychomotor vigilance and subjective alertness. Sleep. **22**, 171–179 (1999). doi: 10.1093/sleep/22.2.171
38. Lamond, N., Dorrian, J., Roach, G.D., McCulloch, K., Holmes, A.L., Burgess, H.J., Fletcher, A., Dawson, D.: The impact of a week of simulated night work on sleep, circadian phase, and performance. Occupational and Environmental Medicine. **60**, 1–9 (2003). doi: 10.1136/oem.60.11.e13
39. Davis, C.M., Roma, P.G., Hienz, R.D.: A rodent model of the human psychomotor vigilance test: Performance comparisons. Journal of Neuroscience Methods. **259**, 57–71 (2016). doi: 10.1016/j.jneumeth.2015.11.014
40. Basner, M., Dinges, D.F.: Maximizing sensitivity of the Psychomotor Vigilance Test (PVT) to sleep loss. Sleep. **34**, 581–591 (2011). doi: 10.1093/sleep/34.5.581
41. Van Dongen, H.P., Dinges, D.F.: Sleep, circadian rhythms, and psychomotor vigilance. Clinics in Sports Medicine. **24**, 237–249 (2005). doi: 10.1016/j.csm.2004.12.007
42. Basner, M., Mollicone, D.J., Dinges, D.F.: Validity and sensitivity of a brief Psychomotor Vigilance Test (PVT-B) to total and partial sleep deprivation. Acta Astronautica. **69**, 949–959 (2011). doi: 10.1016/j.actaastro.2011.07.015
43. Loh, S., Lamond, N., Dorrian, J., Roach, G., Dawson, D.: The validity of psychomotor vigilance tasks of less than 10-minute duration. Behavior Research Methods, Instruments, & Computers. **36**, 339–346 (2004). doi: 10.3758/BF03195580
44. Lamond, N., Dawson, D., Roach, G.D.: Fatigue assessment in the field: Validation of a hand-held electronic psychomotor vigilance task. Aviation, Space, and Environmental Medicine. **76**, 486–489 (2005).
45. Roach, G.D., Dawson, D., Lamond, N.: Can a shorter psychomotor vigilance task be used as a reasonable substitute for the ten-minute psychomotor vigilance task? Chronobiology International. **23**, 379–387 (2006). doi: 10.1080/07420520601067931
46. Dorrian, J., Roach, G.D., Fletcher, A., Dawson, D.: Simulated train driving: Fatigue, self-awareness and cognitive disengagement. Applied Ergonomics. **38**, 155–166 (2007). doi: 10.1016/j.apergo.2006.03.006
47. Lamond, N., Jay, S.M., Dorrian, J., Ferguson, S.A., Roach, G.D., Dawson, D.: The sensitivity of a palm-based psychomotor vigilance task to severe sleep loss. Behavior Research Methods. **40**, 347–352 (2008). doi: 10.3758/BRM.40.1.347
48. Dorrian, J., Rogers, N.L., Dinges, D.F.: Psychomotor vigilance performance: Neurocognitive assay sensitive to sleep loss. In: Kushida, C.A. (eds.) Sleep deprivation: clinical Issues. pharmacology and sleep loss effects. pp. 39–70. Marcel Dekker Inc., New York (2005).

49. Basner, M., Mcguire, S., Goel, N., Rao, H., Dinges, D.F.: A new likelihood ratio metric for the psychomotor vigilance test and its sensitivity to sleep loss. Journal of Sleep Research. **24**, 702–713 (2015). doi: 10.1111/jsr.12322
50. Dinges, D.F., Kribbs, N.B.: Performing while sleepy: Effects of experimentally-induced sleepiness. In: Monk, T.H. (eds.) Sleep, Sleepiness and Performance. pp. 97–128. John Wiley and Sons, Ltd., Chichester (1991).
51. Warm, J.S., Parasuraman, R., Matthews, G.: Vigilance requires hard mental work and is stressful. Human Factors. **50**, 43–441 (2008). doi: 10.1518/001872008X312152
52. Khitrov, M.Y., Laxminarayan, S., Thorsley, D., Ramakrishnan, S., Rajaraman, S., Wesensten, N.J., Reifman, J.: PC-PVT: A platform for psychomotor vigilance task testing, analysis, and prediction. Behavior Research Methods. **46**, 140–147 (2014). doi: 10.3758/s13428-013-0339-9
53. World Medical Association: World Medical Association Declaration of Helsinki Ethical Principles for Medical Research Involving Human Subjects. Jama-Journal of the American Medical Association. **310**, 2191–2194 (2013). doi: 10.1001/jama.2013.281053
54. Kribbs, N.B., Dinges, D.F.: Vigilance decrement and sleepiness. In: Ogilvie, R.D., Harsh, J. (eds.) Sleep Onset: Normal and abnormal processes. pp. 113–125. American Psychological Association, Washington, D.C. (1994).
55. Cohen, J.: Eta-squared and partial eta-squared in fixed factor ANOVA designs. Educational and Psychological Measurement. **33**, 107–112 (1973). doi: 10.1177/001316447303300111
56. Cohen, J.: Statistical power analysis for the behavioral sciences. Erlbaum, Hillsdale, NJ (1988).
57. Rovai, A.P., Baker, J.D., Ponton, M.K.: Social science research design and statistics: A practitioner's guide to research methods and IMB SPS analysis. Watertree Press LLC, Chesapeake, VA (2014).
58. Dorrian, J., Hussey, F., Dawson, D.: Train driving efficiency and safety: Examining the cost of fatigue. Journal of Sleep Research. **16**, 1–11 (2007). doi: 10.1111/j.1365-2869.2007.00563.x
59. Dorrian, J., Baulk, S.D., Dawson, D.: Work hours, workload, sleep and fatigue in Australian Rail Industry employees. Applied Ergonomics. **42**, 202–209 (2011). doi: 10.1016/j.apergo.2010.06.009
60. Dinges, D.F., Powell, J.W.: Sleepiness is more than lapsing. Journal of Sleep Research. **17**, 84 (1988).
61. Dinges, D.F., Powell, J.W.: Sleepiness impairs optimum response capability: It's time to move beyond the lapse hypothesis. Journal of Sleep Research. **18**, 366 (1989).

References

Adan, A. (2013). Cognitive performance and dehydration. *Journal of the American College of Nutrition*, 31, 71–78. doi: 10.1080/07315724.2012.10720011

Adan, A., Archer, S. N., Hidalgo, M. P., Di Milia, L., Natale, V., & Randler, C. (2012). Circadian Typology: A Comprehensive Review. *Chronobiology International*, 29, 1153–1175. doi: 10.3109/07420528.2012.719971

Adlam, A. L. R., Patterson, K., Bozeat, S., & Hodges, J. R. (2010). The Cambridge Semantic Memory Test Battery: Detection of semantic deficits in semantic dementia and Alzheimer's disease. *Neurocase*, 16, 193–207. doi: 10.1080/13554790903405693

Affrunti, N. W., Mehta, T., Rusch, D., & Frazier, S. (2018). Job demands, resources, and stress among staff in after school programs: Neighborhood characteristics influence associations in the job demands resources model. *Children and Youth Services Review*, 88, 366–374. doi: 10.1016/j.childyouth.2018.03.031

Ahsberg, E., Kecklund, G., Akerstedt, T., & Gamberale, F. (2000). Shiftwork and different dimensions of fatigue. *International Journal of Industrial Ergonomics*, 26, 457–465. doi: 10.1016/S0169-8141(00)00007-X

Åkerstedt, T., & Ficca, G. (1997). Alertness–enhancing drugs as a countermeasure to fatigue in irregular work hours. *Chronobiology International*, 14, 145–158. doi: 10.3109/07420529709001152

Åkerstedt, T., Fredlund, P., Gillberg, M., & Jansson, B. (2002). Work load and work hours in relation to disturbed sleep and fatigue in a large representative sample. *Journal of Psychosomatic Research*, 53, 585–588. doi: 10.1016/S0022-3999(02)00447-6

Åkerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, 52, 29–37. dio: 10.3109/00207459008994241

Åkerstedt, T., & Landstrom, U. (1998). Work place countermeasures of night shift fatigue. *International Journal of Industrial Ergonomics*, 21, 167–178. 10.1016/S0169-8141(97)00045-0

Åkerstedt, T., & Wright, K. P. (2009). Sleep loss and fatigue in shift work and shift work disorder. *Sleep Medicine Clinics*, 4, 257–271. doi: 10.1016/j.jsmc.2009.03.001

Akhtar, M. J., & Utne, I. B. (2014). Human fatigue's effect on the risk of maritime groundings – A Bayesian network modeling approach. *Safety Science*, 62, 427–440. doi: 10.1016/j.ssci.2013.10.002

Anderson, C., Grunstein, R. R., & Rajaratnam, S. M. W. (2013). Hours of work and rest in the rail industry. *Internal Medicine Journal*, 6, 717–721. doi: 10.1111/imj.1215

Anund, A., Ihlstrom, J., Fors, C., Kecklund, G., & Filtness, A. (2016). Factors associated with self–reported driver sleepiness and incidents in city bus drivers. *Industrial Health*, 54, 337–346.

Arnedt, J. T., Owens, J., Crouch, M., Stahl, J., & Carskadon, M. A. (2005). Neurobehavioral performance of residents after heavy night call vs after alcohol ingestion. *Jama-Journal of The American Medical Association*, 294, 1025–1033. doi: 10.1001/jama.294.9.1025

Arrowvale. (2018, July 20). On Train Monitoring Recorder. *Arrowvale Electronics*. Retrieved from http://www.otmr.com/products/TrainEventRecorder.pdf

Arriva Trains Wales (ATW). (2016). *Working arrangements for drivers 2016*. Unpublished internal document, Arriva Trains Wales (ATW).

Arriva Trains Wales (ATW). (2017a). *Company Procedure: Management of Fatigue*. Unpublished internal document, Arriva Trains Wales (ATW).

Arriva Trains Wales (ATW). (2017b). *Professional driving handbook*. Cardiff: Arriva Trains Wales (ATW).

Arriva Trains Wales (ATW). (2017c). *Your handbook: A guide to terms and conditions of employment*. Cardiff: Arriva Trains Wales (ATW).

Associated Society of Locomotive Steam Enginemen and Firemen (ASLEF). (2019). On Track with Diversity 2019 edition. ASLEF website: https://www.aslef.org.uk

Australian Transport Council (ATC). (2018, October 15). National road safety strategy 2011-2020. Retrieved from https://roadsafety.gov.au/nrss/files/NRSS_2011_2020.pdf

Australian Transport Safety Bureau. (2011). *Independent investigation into the grounding of the Chinese registered bulk carrier Shen Neng 1 on Douglas Shoal, Queensland, on 3 April 2010*. Australian Government ATSB Transport Safety Report website: http://www.atsb.gov.au/media/2478832/mo2010003.pdf#page=0

Baddeley, A. D. (1968). A three–minute reasoning test based on grammatical trans–formation. *Psychonomic Science*, 10, 341–342.

Baddeley, A. D. (1981). The cognitive psychology of everyday life. *British Journal of Psychology*, 72, 257–269. doi: 10.1111/j.2044-8295.1981.tb02184.x

Balfe, N., Sharples, S., & Wilson, J. R. (2018). Understanding is key: An analysis of factors pertaining to trust in a real–world automation system. *Human Factors*, 60, 477–495. doi: 10.1177/0018720818761256

Banks, S., & Dinges, D. F. (2007). Behavioral and physiological consequences of sleep restriction. *Journal of Clinical Sleep Medicine*, 3, 519–528.

Barcikowski, R. S., & Robey, R. R. (1984). Decisions in single group repeated measures analysis: statistical tests and three computer packages. *American Statistician*, 38, 148–150.

Barger, L. K., Rajaratnam, S. M. W., Cannon, C. P., Lukas, M. A., Im, K., Goodrich, E. L., Czeisler, C. A., & O'Donoghue, M. L. (2017). Short sleep duration, obstructive sleep apnea, shiftwork, and the risk of adverse cardiovascular events in patients after an acute coronary syndrome. *Journal of The American Heart Association,* 6, Article number: e006959. doi: 10.1161/JAHA.117.006959

Barnhoorn, J., Haasnoot, E., Bocanegra, B. R., & van Steenbergen, H. (2014). QRTEngine: An easy solution for running online reaction time experiments using Qualtrics. *Behavior Research Methods,* 47, 918–929. doi: 10.3758/s13428-014-0530-7

Basner, M., & Dinges, D. F. (2011). Maximizing sensitivity of the Psychomotor Vigilance Test (PVT) to sleep loss. *Sleep*, 34, 581–591. doi: 10.1093/sleep/34.5.581

Basner, M., Mcguire, S., Goel, N., Rao, H., & Dinges, D. F. (2015). A new likelihood ratio metric for the psychomotor vigilance test and its sensitivity to sleep loss. *Journal of Sleep Research*, 24, 702–713. doi: 10.1111/jsr.12322

Basner, M., Mollicone, D. J., & Dinges, D. F. (2011). Validity and sensitivity of a brief Psychomotor Vigilance Test (PVT–B) to total and partial sleep deprivation. *Acta Astronautica*, 69, 949–959. doi: 10.1016/j.actaastro.2011.07.015

Baulk, S. D., Biggs, S. N., Reid, K. J., van den Heuvel, C. J., & Dawson, D. (2008). Chasing the silver bullet: Measuring driver fatigue using simple and complex tasks. *Accident Analysis and Prevention*, 40, 396–402. doi: 10.1016/j.aap.2007.07.008

Baulk, S. D., Kandelaars, K. J., Lamond, N., Roach, G. D., Dawson, D., & Fletcher, A. (2007). Does variation in workload affect fatigue in a regular 12-hour shift system? *Sleep and Biological Rhythms*, 5, 74–77. doi: 10.1111/j.1479-8425.2006.00249.x

Beaumont, A., Burton, A. R., Lemon, J., Bennett, B. K., Lloyd, A., & Vollmer-Conna, U. (2012). Reduced cardiac vagal modulation impacts on cognitive performance in chronic fatigue syndrome. *PLoS One*, 7, e49518. doi: 10.1371/journal.pone.0049518

Beers, T. M. (2000). Flexible schedules and shift work: Replacing the ‘9-to-5’ workday? *Monthly Labor Review*, 123, 33–40.

Belenky, G., Wesensten, N. J., Thorne, D. R., Thomas, M. L., Sing, H. C., Redmond, D. P., Russo, M. B., & Balkin, T. J. (2003). Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: A sleep dose–response study. *Journal of Sleep Research*, 12, 1–12. doi: 10.1046/j.1365-2869.2003.00337.x

Bell, J. (2008). *Evaluation of the UK rail sector initial fatigue and risk index thresholds: Identifying good practice RSU/08/03*. Derbyshire: Health and Safety Executive.

Bennett, S. A. (2003). Flight crew stress and fatigue in low-cost commercial air operations - an appraisal. International Journal of Risk Assessment and Management, 4, 207–231.

Bergasa, L. M., Nuevo, J., Sotelo, M. A., Barea, R., & Lopez, M. E. (2006). Real-time system for monitoring driver vigilance. *IEEE Transactions on Intelligent Transportation Systems*, 7, 63–77. doi: 10.1109/TITS.2006.869598

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., Olmstead, R. E., Tremoulet, P. D., & Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation Space and Environmental Medicine*, 78, B231–B244.

Besikci, E. B., Tavacioglu, L., & Arslan, O. (2016). The subjective measurement of seafarers' fatigue levels and mental symptoms. *Maritime Policy & Management*, 43, 329–343. doi: 10.1080/03088839.2015.1047426

Beutel, M. E., Junger, C., Klein, E. M., Wild, P., Lackner, K., Blettner, M., Binder, H., Michal, M., Wiltink, J., Brahler, E., & Munzel, T. (2016). Noise annoyance is associated with depression and anxiety in the general population – The contribution of aircraft noise. *PLoS One*, 11, e0155357. doi: 10.1371/journal.pone.0155357

Blanco, J. A., Johnson, M. K., Jaquess, K. J., Oh, H., Lo, L. C., Gentili, R. J., & Hatfield, B. D. (2018). Quantifying cognitive workload in simulated flight using passive, dry EEG measurements. *IEEE Transactions on Cognitive and Developmental Systems*, 10, 373–383. doi: 10.1109/TCDS.2016.2628702

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 1, 307–310.

Bonsall-Clarke, K. (2012). *Operations and Management – Non–technical skills required in train driver role: Developing an integrated approach to NTS training and investment*. London: Rail Safety and Standards Board LTD.

Booker, L. A., Magee, M., Rajaratnam, S. M. W., Sletten, T. L., & Howard, M. E. (2018). Individual vulnerability to insomnia, excessive sleepiness and shift work disorder amongst healthcare shift workers. A systematic review. *Sleep Medicine Reviews*, 41, 220–233. doi: 10.1016/j.smrv.2018.03.005

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental work-load, fatigue and drowsiness. *Neuroscience and Biobehavioral Reviews*, 44, 58–75. doi: 10.1016/j.neubiorev.2012.10.003

Bouret, S., & Sara, S. J. (2002). Locus coeruleus activation modulates firing rate and temporal organization of odour–induced single–cell responses in rat piriform cortex. *European Journal of Neuroscience*, 16, 2371–2382. doi: 10.1046/j.1460–9568.2002.02413.x

Bowler, N. & Gibson, W. H. (2015). *Fatigue and its Contribution to Railway Incidents RSSB Special Topic Report*. London: RSSB.

Bowling, A. (2005). Just one question: If one question works, why ask several? *Journal of Epidemiology and Community Health*, 59, 342–345. doi: 10.1136/jech.2004.021204

Bowman, A. (2015). An illusion of success: The consequences of British rail privatisation. *Accounting Forum*, 39, 51–63.

Brice, C., & Smith, A. (2001). The effects of caffeine on simulated driving, subjective alertness and sustained attention. *Human Psychopharmacology – Clinical and Experimental*, 16, 523–531. doi: 10.1002/hup.327

Brick, J. (2006). Standardization of alcohol calculations in research. *Alcoholism: Clinical and Experimental Research*, 30, 1276–1287.

British Psychological Society (BPS). (2014). *The British Psychological Society: Code of human research ethics*. Leicester: The British Psychological Society.

Brookhuis, K. A., & De Waard, D. (2001). Assessment of drivers' workload: Performance and subjective and physiological indexes. In P.A. Hancock., & P.A. Desmond (Ed.), *Stress, Workload and Fatigue* (pp. 321–333). New Jersey: Lawrence Erlbaum Associates.

Brown, B. W., Lucero, R. J., & Foss, A. B. (1962). A situation where the pearson correlation-coefficient leads to erroneous assessment of reliability. *Journal of Clinical Psychology*, 18, 95–97. doi: 10.1002/1097-4679(196201)18:1<95::AID-JCLP2270180131>3.0.CO;2-2

Brown, J. G., Wieroney, M., Blair, L., & Zhu, S. J. (2014). Measuring subjective sleepiness at work in hospital nurses: validation of a modified delivery format of the Karolinska Sleepiness Scale. *Sleep and Breathing*, 18, 731–739. doi: 10.1007/s11325–013–0935–z

Brunet, JF., Dagenais, D., Therrien, M., Gartenberg, D., & Forest, G. (2017). Validation of sleep-2-Peak: A smartphone application that can detect fatigue-related changes in reaction times during sleep deprivation. *Behavior Research Methods*, 49, 1460–1469. doi: 10.3758/s13428-016-0802-5

Bruton, A., Conway, J. H., & Holgate, S. T. (2000). Reliability: What is it, and how is it measured? *Physiotherapy*, 86, 94–99. doi: 10.1016/S0031-9406(05)61211-4

Bültmann, U., Kant, I., Kasl, S. V., Beurskens, A. J., & van den Brandt, P. A. (2002). Fatigue and psychological distress in the working population psychometrics, prevalence, and correlates. *Journal of Psychosomatic Research*, 52, 445–452.

Butcher, L. (2016). *Rail regulation*. London: House of Commons Library.

Byrne, A. J., Oliver, M., Bodger, O., Barnett, W. A., Williams, D., Jones, H., & Murphy, A. (2010). Novel method of measuring the mental workload of anaesthetists during clinical practice. *British Journal of Anaesthesia*, 105, 767–771. doi: 10.1093/bja/aeq240

Cabras, C., & Mondo, M. (2018). Coping strategies, optimism, and life satisfaction among first-year university students in Italy: Gender and age differences. *Higher Education*, 75, 643–654. doi: 10.1007/s10734-017-0161-x

Caldwell, J. A. (2005). Fatigue in aviation. *Travel Medicine and Infectious Disease*, 3, 85–96. doi: 10.1016/j.tmaid.2004.07.008

Caldwell, J. A., Caldwell, J. L., Thompson, L. A., & Lieberman, H. R. (2019). Fatigue and its management in the workplace. *Neuroscience and Biobehavioral Reviews*, 96, 272–289. doi: 10.1016/j.neubiorev.2018.10.024

Caldwell, J. A., & Gilreath, S. R. (2002). A survey of aircrew fatigue in a sample of US Army aviation personnel. *Aviation Space and Environmental Medicine*, 73, 472–480.

Canadian Council of Motor Transport Administrator (CCMTA). (2010). *Canadian fatigue related collisions: Fatality estimates 2000–2005*. Canadian Council of Motor Transport Administrator website: https://www.ccmta.ca/en/

Capuron, L., Welberg, L., Heim, C., Wagner, D., Solomon, L., Papanicolaou, D. A., Craddock, R. C., Miller, A. H., & Reeves, W. C. (2006). Cognitive dysfunction relates to subjective report of mental fatigue in patients with chronic fatigue syndrome. *Neuropsychopharmacology*, 31, 1777–1784. doi: 10.1038/sj.npp.1301005

Caramia, P., Lauro, G., Pagano, M., & Natale, P. (2017). Automatic Train Operation Systems: a survey on algorithm and performance index. *AEIT International Annual Conference (AEIT)*, Cagliari, Italy, 2017. New York, NY: IEEE.

Carrier, J., & Monk, T. H. (2000). Circadian rhythms of performance: New trends. *Chronobiology International*, 17, 719–732. doi: 10.1081/CBI-100102108

Carskadon, M. A., & Dement, W. C. (1981). Cumulative effects of sleep restriction on daytime sleepiness. *Psychophysiology*, 18, 107–113. doi: 10.1111/j.1469-8986.1981.tb02921.x

Centofanti, S., Banks, S., Colella, A., Dingle, C., Devine, L., Galindo, H., Pantelios, S., Brkic, G., & Dorrian, J. (2018). Coping with shift work-related circadian disruption: A mixed-methods case study on napping and caffeine use in Australian nurses and midwives. *Chronobiology International*, 35, 853–864. doi: 10.1080/07420528.2018.1466798

Charles, L. E., Burchfiel, C. M., Fekedulegn, D., Vila, B., Hartley, T. A., Slaven, J., Mnatsakanova, A., & Violanti, J. M. (2007). Shift work and sleep: The Buffalo police health study. *An International Journal of Police Strategies & Management*, 30, 215–227. doi: 10.1108/13693951071075322 5

Chang, H. L., & Ju, L. S. (2008). Effect of consecutive driving on accident risk: A comparison between passenger and freight train driving. ACCIDENT *Analysis and Prevention*, 40, 1844–1849. doi: 10.1016/j.aap.2008.07.009

Chang, W. P. (2018). Influence of shift type on sleep quality of female nurses working monthly rotating shifts with cortisol awakening response as mediating variable. *Chronobiology International*, 35, 1503–1512. doi: 10.1080/07420528.2018.1493491

Chetty, K., Chen, Q. C., & Woodbridge, K. (2016). Train monitoring using GSM-R based passive radar. In: IEEE (Eds.), *IEEE Radar Conference* (pp. 34–37). Philadelphia, PA.

Ciaramelli, C., Palmioli, A., & Airoldi, C. (2019). Coffee variety, origin and extraction procedure: Implications for coffee beneficial effects on human health. *Food Chemistry*, 278, 47–55. doi: 10.1016/j.foodchem.2018.11.063

Ciombor, K. J., Ennis, M., & Shipley, M. T. (1999). Norepinephrine increases rat mitral cell excitatory responses to weak olfactory nerve input via alpha-1 receptors in vitro. *Neuroscience*, 90, 595–606. doi: 10.1016/S0306–4522(98)00437–0

Clancy, J. (2007). Assessing and managing human factors risks: Practical examples in the Australian rail. In: M. J. Anca Jr. (Eds.), *Multimodal safety management and human factor: Crossing the borders of medical, aviation, road and rail industry*. Boca Raton, FL: Taylor & Francis Group.

Clark, W. W., & Popelka, G. R. (1989). Hearing levels of railroad trainmen. *Laryngoscope*, 99, 1151–1157.

Cohen, J. (1973). Eta–squared and partial eta–squared in fixed factor ANOVA designs. *Educational and Psychological Measurement,* 33, 107–112. doi: 10.1177/001316447303300111

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological bulletin*, 112, 155–159.

Comstock, J. R. J., & Arnegard, R. J. (1992). *The Multi-Attribute Test Battery for Human Operator Workload and Strategic Behaviour*. Hampton: National Aeronautics and Space Administration.

Connor, J., Norton, R., Ameratunga, S., Robinson, E., Civil, I., Dunn, R., Bailey, J., & Jackson, R. (2002). Driver sleepiness and risk of serious injury to car occupants: Population based case control study. *British Medical Journal*, 324, 1125–1128A. doi: 10.1136/bmj.324.7346.1125

Conway, P., Campanini, P., Sartori, S., Dotti, R., & Costa, G. (2008). Main and interactive effects of shiftwork, age and work stress on health in an Italian sample of healthcare workers. *Applied Ergonomics*, 39, 630–639. doi: 10.1016/j.apergo.2008.01.007

Costa, G., & Sartori, S. (2007). Ageing, working hours and work ability. *Ergonomics*, 50, 1914–1930. doi: 10.1080/00140130701676054

Costigan, A., Elliott, C., McDonald, C., & Newton, J. L. (2010). Orthostatic symptoms predict functional capacity in chronic fatigue syndrome: implications for management. QJM-*An International Journal of Medicine*, 103, 589–595. doi: 10.1093/qjmed/hcq094

Cotrim, T., Carvalhais, J., Neto, C., Teles, N., Noriega, P., & Rebelo, F. (2017). Determinants of sleepiness at work among railway control workers. *Applied Ergonomics*, 58, 293–300. doi: 10.1016/j.apergo.2016.07.006

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS One*, 8, e57410. doi:10.1371/journal. pone.0057410

Cushway, D., & Tyler, P. A. (1994). Stress and coping in clinical psychologists. *Stress Medicine*, 10, 35–42. doi: 10.1002/smi.2460100107

Dahl, R. E., & Lewin, D. S. (2002). Pathways to adolescent health: Sleep regulation and behavior. *Journal of Adolescent Health*, 31, 175–184. doi: 10.1016/S1054-139X(02)00506-2

Daly, L. E., & Bourke, G. J. (2000). *Interpretation and use of medical statistics*. Oxford: Blackwell Science Ltd.

Darwent, D., Dawson, D., Paterson, J. L., Roach, G. D., & Ferguson, S. A. (2015). Managing fatigue: It really is about sleep. *Accident Analysis and Prevention*, 82, 20–26. doi: 10.1016/j.aap.2015.05.009

Darwent, D., Lamond, N., & Dawson, D. (2008). The sleep and performance of train drivers during an extended freight-haul operation. *Applied Ergonomics*, 5, 614–622. doi: 10.1016/j.apergo.2008.02.012

Darbra, R. M., & Casal, J. (2004). Historical analysis of accidents in seaports. *Safety Science*, 42, 85–98. doi: 10.1016/S0925-7535(03)00002-X

Davis, S., Mirick, D. K., & Stevens, R. G. (2001). Night shift work, light at night, and risk of breast cancer. *JNCI-Journal of The National Cancer Institute*, 93, 1557–1562. doi: 10.1093/jnci/93.20.1557

Davis, C. M., Roma, P. G., & Hienz, R. D. (2016). A rodent model of the human psychomotor vigilance test: Performance comparisons. *Journal of Neuroscience Methods*, 259, 57–71. doi: 10.1016/j.jneumeth.2015.11.014

Dawson, A., & Reid, K. (1997). Fatigue, alcohol, and performance impairment. *Nature,* 388, 235. doi: 10.1038/40775

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

Dawson, D., Searle, A. K., & Paterson, J. L. (2014). Look before you (s)leep: Evaluating the use of fatigue detection technologies within a fatigue risk management system for the road transport industry. Sleep Medicine Reviews, 18, 141–152.

De Boer, A. G. E. M., van Lanschot, J. J. B., Stalmeier, P. F. M., van Sandick, J. W., Hulscher, J. B. F., De Haes, J. C. J. M., & Sprangers, M. A. G. (2004). Is a single-item visual analogue scale as valid, reliable and responsive as multi-item scales in measuring quality of life? *Quality of Life Research*, 13, 311–320. doi: 10.1023/B:QURE.0000018499.64574.1f

de Godoy, L. B. M., Luz, G. P., Palombini, L. O., Silva, L. O. E., Hoshino, W., Guimaraes, T. M., Tufik, S., Bittencourt, L., & Togeiro, S. M. (2016). Upper airway resistance syndrome patients have worse sleep quality compared to mild obstructive sleep apnea. *PLoS One*, 11, Article number: e0156244. doi: 10.1371/journal.pone.0156244

de Leeuw, J. R., & Motz, B. A. (2015). Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavior Research Methods*, 1–12. doi: 10.3758/s13428-015-0567-2

Dekker, D. K., Paley, M. J., Popkin, S. M., & Tepas, D. I. (1993). Locomotive engineers and their spouses - coffee consumption, mood, and sleep reports. *Ergonomics*, 36, 233–238. doi: 10.1080/00140139308967877

DeVon, H. A., Block, M. E., Moyle-Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J., Savoy, S. M., & Kostas-Polston, E. (2007). A psychometric toolbox for testing validity and reliability. *Journal of Nursing Scholarship*, 39, 155–164. doi: 10.1111/j.1547-5069.2007.00161.x

Dinges, D. F., & Kribbs, N. B. (1991). Performing while sleepy: Effects of experimentally-induced sleepiness. In: T. H. Monk (Ed.), *Sleep, sleepiness and performance* (pp. 97–128). Chichester: John Wiley and Sons, Ltd.

Dinges, D. F., Orne, M. T., Whitehouse, W. G., & Orne, E. C. (1987). Temporal placement of a nap for alertness: contributions of circadian phase and prior wakefulness. *Sleep*, 10, 313–329.

Dinges, D. F., Pack, F., Williams, K., Gillen, K. A., Powell, J. W., Ott, G. E., Aptowicz, C., & Pack, A. I. (1997). Cumulative sleepiness, mood disorder, and psychomotor vigilance performance decrements during a week of sleep restricted to 4–5 hours per night. *Sleep*, 20, 267-277.

Dinges, D. F., & Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments, & Computers*, 17, 652–655.

Dinges, D. F., & Powell, J. W. (1988). Sleepiness is more than lapsing. *Sleep Research*, 17, 84.

Dinges, D. F. & Powell, J. W. (1989). Sleepiness impairs optimum response capability. *Sleep Research*, 18, 366.

Dixon-Woods, M., & Bosk, C. (2010). Learning through observation: the role of ethnography in improving critical care. *Current Opinion in Critical Care*, 16, 639–642. doi: 10.1097/MCC.0b013e32833ef5ef

Doerr, J. M., Ditzen, B., Strahler, J., Linnemann, A., Ziemek, J., Skoluda, N., Hoppmann, C. A., & Nater, U. M. (2015). Reciprocal relationship between acute stress and acute fatigue in everyday life in a sample of university students. *Biological Psychology*, 110, 42-49. doi: 10.1016/j.biopsycho.2015.06.009

Dong, H. R., Ning, B., Cai, B. G., & Hou, Z. S. (2010). Automatic train control system development and simulation for high–speed railways. *IEEE Circuits and Systems Magazine*, 10, 6–18. doi: 10.1109/MCAS.2010.936782

Doran, S. M., van Dongen, H. P. A., & Dinges, D. F. (2001). Sustained attention performance during sleep deprivation: Evidence of state instability. *Archives Italiennes de Biologie*, 139, 253–267.

Dorrian, J., Baulk, S. D., & Dawson, D. (2011). Work hours, workload, sleep and fatigue in Australian rail industry employees. *Applied Ergonomics*, 42, 202–209. doi: 10.1016/j.apergo.2010.06.009

Dorrian, J., Hussey, F., & Dawson, D. (2007). Train driving efficiency and safety: examining the cost of fatigue. *Journal of Sleep Research*, 16, 1–11. doi: 10.1111/j.1365-2869.2007.00563.x

Dorrian, J., Lamond, N., & Dawson, D. (2000). The ability to self–monitor performance when fatigued. *Journal of Sleep Research*, 9, 137–145.

Dorrian, J., Lamond, N., Kozuchowski, K., & Dawson, D. (2008). The driver vigilance telemetric control system (DVTCS): Investigating sensitivity to experimentally induced sleep loss and fatigue. *Behavior Research Methods*, 40, 1016–1025. doi: 10.3758/BRM.40.4.1016

Dorrian, J., Lamond, N., Holmes, A. L., Burgess, H. J., Roach, G. D., Fletcher, A., & Dawson, D. (2003). The ability to self-monitor performance during a week of simulated night shifts. *Sleep*, 26, 871–877. doi: 10.1093/sleep/26.7.871

Dorrian, J., Roach, G. D., Fletcher, A., & Dawson, D. (2007). Simulated train driving: Fatigue, self-awareness and cognitive disengagement. *Applied Ergonomics*, 38, 155–166. dio: 10.1016/j.apergo.2006.03.006

Dorrian, J., Rogers, N. L., & Dinges, D. F. (2005). Psychomotor vigilance performance: Neurocognitive assay sensitive to sleep loss. In: C. A. Kushida (Ed.), *Sleep deprivation: Clinical issues. Pharmacology and sleep loss effects* (pp. 39–70). New York, NY: Marcel Dekker Inc.

Dragone, D. (2009). I am getting tired: Effort and fatigue in intertemporal decision-making. *Journal of Economic Psychology*, 30, 552–562. doi: 10.1016/j.joep.2009.03.008

Driskell, J. E., & Mullen, B. (2005). The efficacy of naps as a fatigue countermeasure: A meta-analytic integration. *Human Factors*, 47, 360–377. doi: 10.1518/0018720054679498

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. doi: 10.1080/00140139.2012.691994

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

Edworthy, J., Reid, S., Peel, K., Lock, S., Williams, J., Newbury, C., Foster, J., & Farrington, M. (2018). The impact of workload on the ability to localize audible alarms. *Applied Ergonomics*, 72, 88–93. doi: 10.1016/j.apergo.2018.05.006

Eggemeier, F. T., Wilson, G. F., Kramer, A. F., & Damos, D.L. (1991). General considerations concerning workload assessment in multi-task environments. In: D. L. Damos (Eds.), *Multiple task performance* (pp. 207–216). London: Taylor & Francis.

Einother, S. J. L., & Giesbrecht, T. (2013). Caffeine as an attention enhancer: Reviewing existing assumptions. P*sychopharmacology*, 225, 251–274. doi: 10.1007/s00213-012-2917-4

Elmenhorst, E. M., Quehl, J., Müller, U., & Basner, M. (2014). Nocturnal air, road, and rail traffic noise and daytime cognitive performance and annoyance. *Journal of the Acoustical Society of America*, 135, 213–222. doi: 10.1121/1.4842475

Evans, M. S., Harborne, D., & Smith, A. P. (2019). Exploring an objective indicator of fatigue: A modified mobile version of the Psychomotor Vigilance Task (m-PVT). In: L. Longo and M.C. Leva (Eds.), *Human Mental Workload: Models and Applications. H–WORKLOAD 2018. Communications in Computer and Information Science* (pp. 49–71). Cham: Springer.

Evirgen, N., Oniz, A., & Ozgoren, M. (2015). The development of a psychomotor vigilance test battery and the assesment of vigilance change in reaction to mental work load. *Journal of Neurological Sciences-Turkish*, 32, 494–506.

Fan, J., & Smith, A. P. (2017a). The impact of workload and fatigue on performance. In: L. Longo, & M. C. Leva (Eds.), *Human Mental Workload: Models and Applications. H–WORKLOAD 2017*. Communications in Computer and Information Science (pp. 90–105). Cham: Springer.

Fan, J., & Smith, A. (2017b). Positive well–being and work–life balance among UK railway staff. *Open Journal of Social Sciences*, 5, 1–6. doi: 10.4236/jss.2017.56001

Fan, J. L., & Smith, A. P. (2018). A preliminary review of fatigue among rail staff. *Frontiers in Psychology*, 9, Article number: 634. doi: 10.3389/fpsyg.2018.00634

Fan, J., & Smith, A. P. (2019). Mental workload and other causes of different types of fatigue in rail staff. In: L. Longo, & M. C. Leva (Eds.), *Human Mental Workload: Models and Applications. H–WORKLOAD 2018. Communications in Computer and Information Science* (pp. 147–159). Cham: Springer.

Federal Railroad Administration (FRA). (2015). Office of Safety Analysis website: http://safetydata.fra.dot.gov/OfficeofSafety/publicsite/Query/rrchart.aspx

Feldman, G., & Hayes, A. (2005). Preparing for problems: A measure of mental anticipatory processes. *Journal of Research in Personality*, 39, 487–516. doi: 10.1016/j.jrp.2004.05.005

Ferketich, S. (1990). Focus on psychometrics internal consistency estimates of reliability. *Research in Nursing & Health*, 13, 437–440. doi: 10.1002/nur.4770130612

Fernandes, R., Hatfield, J., & Soames, R. F. (2010). A systematic investigation of the differential predictors for speeding, drink-driving, driving while fatigued, and not wearing a seat belt, among young drivers. *Transportation Research Part F-Traffic Psychology and Behaviour*, 13, 179–196. doi: 10.1016/j.trf.2010.04.007

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th Ed.). London: SAGE Publications Ltd.

Fisher, G. G., Matthews, R. A., & Gibbons, A. M. (2016). Developing and investigating the use of single-item measures in organizational research. *Journal of Occupational Health Psychology*, 21, 3–23. doi: 10.1037/a0039139

Fletcher, A., & Dawson, D. (2001). Field-based validations of a work-related fatigue model based on hours of work. *Transportation Research Part F: Traffic Psychology and Behaviour*, 4, 75–88. doi: 10.1016/S1369-8478(01)00015-8

Flo, E., Pallesen, S., Mageroy, N., Moen, B. E., Gronli, J., Nordhus, I. H., & Bjorvatn, B. (2012). Shift work disorder in nurses: Assessment, prevalence and related health problems. *PLoS One*, 7, e33981. doi: 10.1371/journal.pone.0033981

Folkard, S. (1983). Diurnal variation in human performance. In: R. Hockey (Eds.), *Stress and fatigue in human performance*. New York: John Wiley.

Folkard, S., & Lombardi, D. A. (2006). Modeling the impact of the components of long work hours on injuries and "accidents". *American Journal of Industrial Medicine*, 11, 953–963. doi: 10.1002/ajim.20307

Folkard, S., & Monk, T. H. (1985). Circadian performance rhythms. In: S. Folkard, & T. H. Monk (Eds.), *Hours of work—temporal factors in work scheduling* (pp. 37–52). New York: John Wiley and Sons.

Folkard, S., & Tucker, P. (2003). Shift work, safety and productivity. *Occupational Medicine-Oxford*, 53, 95–101. doi: 10.1093/occmed/kqg047

Fonken, L. K., Workman, J. L., Walton, J. C., Weil, Z. M., Morris, J. S., Haim, A., & Nelson, R. J. (2010). Light at night increases body mass by shifting the time of food intake. *Proceedings of The National Academy of Sciences of The United States of America*, 43, 18664–18669. doi: 10.1073/pnas.1008734107

Foy, H. J., & Chapman, P. (2018). Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation. *Applied Ergonomic*, 73, 90–99. doi: 10.1016/j.apergo.2018.06.006

Fukuda, K., Straus, S. E., Hickie, I., Sharpe, M. C., Dobbins, J. G., Komaroff, A., Schluederberg, A., Jones, J. F., & Lloyd, A. R. (1994). The chronic fatigue syndrome - A comprehensive approach to its definition and study. *Annals of Internal Medicine*, 121, 953–959. 10.7326/0003-4819-121-12-199412150-00009

Fukuda, S., Yamano, E., Joudoi, T., Mizuno, K., Tanaka, M., Kawatani, J., Takano, M., Tomoda, A., Imai-Matsumura, K., Miike, T., & Watanabe, Y. (2010). Effort-reward imbalance for learning is associated with fatigue in school children. *Behavioral Medicine*, 36, 53–62. doi: 10.1080/08964281003774919

Gallagher, D., Heymsfield, S. B., Heo, M., Jebb, S. A. Murgatroyd, P. R., & Sakamoto, Y. (2000). Healthy percentage body fat ranges: an approach for developing guidelines based on body mass index. *American Journal of Clinical Nutrition*, 72, 694–701.

Gander, P. H., Mangie, J., van den Berg, M. J., Smith, A. A. T., Mulrine, H. M., & Signal, TL. (2014a). Crew fatigue safety performance indicators for fatigue risk management systems. *Aviation Space and Environmental Medicine*, 85, 139–147. doi: 10.3357/ASEM.3748.2014

Gander, P. H., Mulrine, H. M., van den Berg, M. J., Smith, A. A. T., Signal, T. L., Wu, L. J., & Belenky, G. (2014b). Pilot fatigue: Relationships with departure and arrival times, flight duration, and direction. *Aviation Space and Environmental Medicine*, 85, 833–840. doi: 10.3357/ASEM.3963.2014

Gander, P. H., Mulrine, H. M., van den Berg, M. J., Smith, A. A. T., Signal, T. L., Wu, L. J., & Belenky, G. (2015). Effects of sleep/wake history and circadian phase on proposed pilot fatigue safety performance Indicators. *Journal of Sleep Research*, 24, 110–119. doi: 10.1111/jsr.12197

Gander, P., Mulrine, H. M., van den Berg, M. J., Wu, L., Smith, A., Signal, L., & Mangie, J. (2016). Does the circadian clock drift when pilots fly multiple transpacific flights with 1-to 2-day layovers? *Chronobiology International*, 33, 982–994. doi: 10.1080/07420528.2016.1189430

Gander, P. H., Signal, T. L., van den Berg, M. J., Mulrine, H. M., Jay, S. M., & Mangie, C. J. (2013). In-flight sleep, pilot fatigue and psychomotor vigilance task performance on ultra-long range versus long range flights. *Journal of Sleep Research*, 22, 697–706. doi: 10.1111/jsr.12071

Ganio, M. S., Armstrong, L. E., Casa, D. J., McDermott, B. P., Lee, E. C., Yamamoto, L. M., Marzano, S., Lopez, R. M., Jimenez, L., Le Bellego, L., Chevillotte, E., & Lieberman, H. R. (2011). *British Journal of Nutrition*, 106, 1535–1543. doi: 10.1017/S0007114511002005

Gillberg, M., Kecklund, G., & Akerstedt, T. (1994). Relations between performance and subjective ratings of sleepiness during a night awake. *Sleep*, 17, 236–241. doi: 10.1093/sleep/17.3.236

Girden, E. R. (1992). *ANOVA: Repeated measures*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-084. Newbury Park, CA: Sage.

Gisolfi, C. V., & Copping, J. R. (1993): Thermal effects of prolonged treadmill exercise in the heat. *Medicine and Science in Sports and Exercise*, 25, 310–315.

Goel, N., Abe, T., Braun, M. E., & Dinges, D. F. (2014). Cognitive workload and sleep restriction interact to influence sleep homeostatic responses. *Sleep*, 37, 1746–1757. doi: 10.5665/sleep.4164

Goode, J. H. (2003). Are pilots at risk of accidents due to fatigue? *Journal of Safety Research*, 34, 309–313. doi: 10.1016/S0022-4375(03)00033-1

Gorlova, S., Ichiba, T., Nishimaru, H., Takamura, Y., Matsumoto, J., Hori, E., Nagashima, Y., Tatsuse, T., Ono, T., & Nishijo, H. (2019). Non-restorative sleep caused by autonomic and electroencephalography parameter dysfunction leads to subjective fatigue at wake rime in shift workers. *Frontiers in Neurology*, 10, Article number: 66. doi: 10.3389/fneur.2019.00066

Grant, S. S., Magruder, K. P., & Friedman, B. H. (2018). Controlling for caffeine in cardiovascular research: A critical review. *International Journal of Psychophysiology*, 133, 193–201. doi: 10.1016/j.ijpsycho.2018.07.001

Grech, M. R., Neal, A., Yeo, G., Humphreys, M., & Smith, S. (2009). An examination of the relationship between workload and fatigue within and across consecutive days of work: Is the relationship static or dynamic? *Journal of Occupational Health Psychology*, 14, 231–242. doi: 10.1037/a0014952

Greene, C. M., Bellgrove, M. A., Gill, M., & Robertson, I. H. (2009). Noradrenergic genotype predicts lapses in sustained attention. *Neuropsychologia*, 47, 591–594. doi: 10.1016/j.neuropsychologia.2008.10.003

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95–112.

Gui, D. Y., Xu, S. H., Zhu, S. H., Fang, Z., Spaeth, A. M., Xin, Y. Y., Feng, T. Y., & Rao, H. Y. (2015). Resting spontaneous activity in the default mode network predicts performance decline during prolonged attention workload. *Neuroimage*, 120, 323–330. doi: 10.1016/j.neuroimage.2015.07.030

Gunzelmann, G., Moore, L. R., Gluck, K. A., van Dongen, H. P. A., & Dinges, D. F. (2011). Fatigue in sustained attention: Generalizing mechanisms for time awake to time on task. In P. L. Ackerman (Ed.), *Decade of behavior/science conference. cognitive fatigue: Multidisciplinary perspectives on current research and future applications* (pp. 83-101). Washington, DC, US: American Psychological Association.

Guo, F., Wang, T. B., & Ning, Z. J. (2017). Subjective measures of work-related fatigue in automobile factory employees. *Work-A Journal of Prevention Assessment & Rehabilitation*, 58, 233–240. doi: 10.3233/WOR-172606

Ha, M., & Park, J. (2005). Shiftwork and metabolic risk factors of cardiovascular disease. *Journal of Occupational Health*, 47, 89–95. doi: 10.1539/joh.47.89

Hancock, P. A., Ross, J. M., & Szalma, J. L. (2007). A meta–analysis of performance response under thermal stressors. *Human Factors*, 49, 851–877.

Hansen, J. (2001). Light at night, shiftwork, and breast cancer risk. *JNCI-Journal of The National Cancer Institute*, 93, 1513–1515. doi: 10.1093/jnci/93.20.1513

Harma, M., Sallinen, M., Ranta, R., Mutanen, P., & Muller, K. (2002). The effect of an irregular shift system on sleepiness at work in train drivers and railway traffic controllers. *Journal of Sleep Research*, 11, 141–151. doi: 10.1046/j.1365-2869.2002.00294.x

Harma, M., Tarja, H., Irja, K., Mikael, S., Jussi, V., Anne, B., & Pertti, M. (2006). A controlled intervention study on the effects of a very rapidly forward rotating shift system on sleep-wakefulness and well-being among young and elderly shift workers. *International Journal of Psychophysiology*, 59, 70–79. doi: 10.1016/j.ijpsycho.2005.08.005

Harrison, S., Smith, A. P., & Sykes, R. (2002). Residential rehabilitation courses in the self-directed management of chronic fatigue syndrome. *Journal of Chronic Fatigue Syndrome*, 10, 59–65. 10.1300/J092v10n02_05

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Advances in Psychology*, 52, 139–183. doi: 10.1016/S0166-4115(08)62386-9

Harvanko, A. M., Derbyshire, K. L., Schreiber, L. R. N., & Grant, J. E. (2015). The effect of self-regulated caffeine use on cognition in young adults. *Human Psychopharmacology: Clinical and Experimental*, 30, 123–130. doi: 10.1002/hup.2464

Haskell, C. F., Kennedy, D. O., Wesnes, K. A., & Scholey, A. B. (2005). Cognitive and mood improvements of caffeine in habitual consumers and habitual non-consumers of caffeine. *Psychopharmacology*, 179, 813–825. doi: 10.1007/s00213-004-2104-3

Hayat, J., & Redfern, R. (2018). *Thameslink programme evaluation: Baseline report Final Report to Department for Transport (DfT)*. London: AECOM Limited.

Health and Safety Executive (HSE). (2006). *Managing shiftwork: Health and safety guidance*. Health and Safety Executive website: http://www.hse.gov.uk/pubns/priced/hsg256.pdf.

Heath, C., Hindmarsh, J., & Luff, P. (1999). Interaction in isolation: The dislocated world of the London Underground train driver. *Sociology – The Journal of the British Sociological Association*, 3, 555–575. doi: 10.1017/S0038038599000358

Heine, T., Lenis, G., Reichensperger, P., Beran, T., Doessel, O., & Deml, B. (2017). Electrocardiographic features for the measurement of drivers' mental workload. *Applied Ergonomics*, 61, 31–43. doi: 10.1016/j.apergo.2016.12.015

Henderson, D., & Saunders, S. S. (1998). Acquisition of noise–induced hearing loss by railway workers. *Ear and Hearing*, 19, 120–130. doi: 10.1097/00003446–199804000–00004

Higdon, J. V., & Frei, B. (2006). Coffee and health: A review of recent human research. *Critical Reviews in Food Science and Nutrition*, 46, 101–123. doi: 10.1080/10408390500400009

Hittle, B. M., & Gillespie, G. L. (2018). Identifying shift worker chronotype: implications for health. *Industrial Health*, 56, 512–523. doi: 10.2486/indhealth.2018-0018

Holy, L. (1984). Theory, methodology and the research process. In: R. Ellen (Ed.), *Ethnographic research: A guide to general conduct* (pp. 13–34). London: Academic Press.

Hogervorst, M. A., Brouwer, A. M., & Van Erp, J. B. E. (2014). Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental work-load. *Frontiers in Neuroscience*, 8, 322. doi: 10.3389/fnins.2014.00322

Honn, K. A., Satterfield, B. C., McCauley, P., Caldwell, J. L., & van Dongen, HPA. (2016). Fatiguing effect of multiple take-offs and landings in regional airline operations. *Accident Analysis and Prevention*, 86, 199–208. doi: 10.1016/j.aap.2015.10.005

Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine*, 30, 1–15. doi: 10.2165/00007256-200030010-00001

Hou, R. H., Moss-Morris, R., Risdale, A., Lynch, J., Jeevaratnam, P., Bradley, B. P., & Mogg, K. (2014). Attention processes in chronic fatigue syndrome: Attentional bias for health-related threat and the role of attentional control. *Behaviour Research and Therapy*, 52, 9–16. doi: 10.1016/j.brat.2013.10.005

Hsieh, M. L., Li, Y. M., Chang, E. T., Lai, H. L., Wang, W. H., & Wang, S. C. (2011). Sleep disorder in Taiwanese nurses: A random sample survey. *Nursing & Health Sciences*, 13, 468–474. doi: 10.1111/j.1442-2018.2011.00641.x

Hsu, B. W., Wang, M. J. J., Chen, C. Y., & Chen, F. (2015). Effective indices for monitoring mental workload while performing multiple tasks. *Perceptual and Motor Skills*, 121, 94–117. doi: 10.2466/22.PMS.121c12x5

Hughes, J. M., Ulmer, C. S., Gierisch, J. M., & Howard, M. O. (2018). Single-item measures for detecting sleep problems in united states military veterans. *Journal of General Internal Medicine*, 33, 698–704. doi: 10.1007/s11606-017-4250-4

Hursh, S. R., Redmond, D. P., Johnson, M. L., Thorne, D. R., Belenky, G., Balkin, T. J., Storm, W. F., Miller, J. C., & Eddy D. R. (2004). Fatigue models for applied research in warfighting. *Aviation, Space, and Environmental Medicine*, 75, A44–A53.

Hursh, S. R., & van Dongen, H. P. A. (2010). Fatigue and performance modeling. In: M. H. Kryger, T. Roth, W. C. Dement (Eds.), *Principles and practice of sleep medicine* (pp. 745–752). St. Louos: Elsevier Saunders.

Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomised block and split-plot designs. *Journal of Educational Statistics*, 1, 69–82.

Iodice, P., Calluso, C., Barca, L., Bertollo, M., Ripari, P., & Pezzulo, G. (2017). Fatigue increases the perception of future effort during decision making. *Psychology of Sport and Exercise*, 33, 150–160. doi: 10.1016/j.psychsport.2017.08.013

Jamal, M. (2004). Burnout, stress and health of employees on non-standard work schedules: A study of Canadian workers. *Stress & Health*, 20, 113–119.

Jay, S. M., Dawson, D., Ferguson, S. A., & Lamond, N. (2008). Driver fatigue during extended rail operations. *Applied Ergonomics*, 39, 623–629. doi: 10.1016/j.apergo.2008.01.011

Jepsen, J. R., Zhao, Z. W., & van Leeuwen, W. M. A. (2015). Seafarer fatigue: a review of risk factors, consequences for seafarers' health and safety and options for mitigation. *International Maritime Health*, 66, 106–117. doi: 10.5603/IMH.2015.0024

Jewett, M. E., Dijk, D. J., Kronauer, R. E., & Dinges, D. F. (1999). Dose–response relationship between sleep duration and human psychomotor vigilance and subjective alertness. *Sleep*, 22, 171–179. doi: 10.1093/sleep/22.2.171

Jimenez-Molina, A., Retamal, C., & Lira, H. (2018). Using psychophysiological sensors to assess mental workload during web browsing. *Sensors*, 18, 458. doi: 10.3390/s18020458

Jobin, J., Wrosch, C., & Scheier, M. F. (2014). Associations between dispositional optimism and diurnal cortisol in a community sample: When stress is perceived as higher than normal. *Health Psychology*, 33, 382–391. doi: 10.1037/a0032736

Kaida, K., Takahashi, M., Åkerstedt, T., Nakata, A., Otsuka, Y., Haratani, T., & Fukasawa, K. (2006). Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clinical Neurophysiology*, 117, 1574–1581.

Karasek, R., & Theorell, T. (1990) *Healthy work: Stress, productivity and the reconstruction of working life*. New York: Basic Books.

Kathner, I., Wriessnegger, S. C., Muller-Putz, G. R., Kubler, A., & Halder, S. (2014). Effects of mental workload and fatigue on the P300, alpha and theta band power during operation of an ERP (P300) brain-computer interface. *Biological Psychology*, 102, 118–129. 10.1016/j.biopsycho.2014.07.014

Kazemi, Z., Mazloumi, A., Saraji, G. N., & Barideh, S. (2016). Fatigue and workload in short and long-haul train driving. *Work-A Journal of Prevention Assessment & Rehabilitation*, 54, 425–433. doi: 10.3233/WOR-162328

Keller, F., Gunasekharan, S., Mayo, N., & Corley, M. (2009). Timing accuracy of web experiments: A case study using the WebExp software package. *Behavior Research Methods*, 41, 1–12. doi: 10. 3758/BRM.41.1.12

Kelley, A. M., Feltman, K. A., & Curry, I. P., (2018). A survey of fatigue in army aviators. *Aerospace Medicine and Human Performance*, 89, 464–468. doi: 10.3357/AMHP.5044.2018

Kenvyn, F. (2007). Fatigue management on the operational railway: Developing a strategic approach. In: J. R. Wilson, B. Norris, T. Clarke, & A. Mills (Eds.), *People and rail systems: Human factors at the heart of the railway* (pp. 535–544). Aldershot: Ashgate Publishing Limited.

Keppel, G. (1991). *Design and Analysis: A researcher's handbook*. Englewood Cliffs, NJ: Prentice Hall.

Kerin, A., & Aguirre, A. (2005). Improving health, safety, and profits in extended hours operations (Shiftwork). *Industrial Health*, 43, 201–208. doi: 10.2486/indhealth.43.201

Khan, S., Duan, P. F., Yao, L. G., & Hou, H. W. (2018). Shiftwork-Mediated Disruptions of Circadian Rhythms and Sleep Homeostasis Cause Serious Health Problems. *International Journal of Genomics*, Article number: 8576890. doi: 10.1155/2018/8576890

Khitrov, M. Y., Laxminarayan, S., Thorsley, D., Ramakrishnan, S., Rajaraman, S., Wesensten, N. J., & Reifman, J. (2014). PC–PVT: A platform for psychomotor vigilance task testing, analysis, and prediction. *Behavior Research Methods*, 46, 140–147. doi: 10.3758/s13428-013-0339-9

Killgore, W. D. S. (2010). Effects of sleep deprivation on cognition. *Human Sleep and Cognition, Part I: Basic Research*, 185, 105–129. doi: 10.1016/B978-0-444-53702-7.00007-5

Kim, H. J., & Abraham, I. (2017). Measurement of fatigue: Comparison of the reliability and validity of single-item and short measures to a comprehensive measure. *International Journal of Nursing Studies*, 65, 35–43. doi: 10.1016/j.ijnurstu.2016.10.012

Knauth, P., & Hornberger, S. (2003). Preventive and compensatory measures for shift workers. *Occupational Medicine-Oxford*, 53, 109–116. doi: 10.1093/occmed/kqg049

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 16, 155–163. doi: 10.1016/j.jcm.2016.02.012

Korunka, C., Kubicek, B., Prem, R., & Cvitan, A. (2012). Recovery and detachment between shifts, and fatigue during a twelve–hour shift. *Work – A Journal of Prevention Assessment & Rehabilitation*, 41, 3227–3233. doi: 10.3233/WOR–2012-0587-3227

Kosmadopoulos, A., Sargent, C., Zhou, X., Darwent, D., Matthews, R. W., Dawson, D., & Roach, G. D. (2017). The efficacy of objective and subjective predictors of driving performance during sleep restriction and circadian misalignment. *Accident Analysis and Prevention*, 99, 445–451. doi: 10.1016/j.aap.2015.10.014

Kosslyn, M. S., & Smith, E. E. (2001). Higher cognitive functions: Introduction. In: M. S. Gassaniga (Eds.). *The new cognitive neurosciences* (pp. 961–964). Cambridge: The MIT Press.

Krausz, M., Sagie, A., & Bidermann, Y. (2000). Actual and preferred work schedules and scheduling control as determinants of job-related attitudes. *Journal of Vocational Behavior*, 56, 1–11.

Kribbs, N. B., & Dinges, D. F. (1994). Vigilance decrement and sleepiness. In: J. Harsh, & R. Ogilvie (Eds.), *Sleep onset mechanisms* (pp. 113–125). Washington: American Psychological Association.

Kroemer, K. E. H., & Grandjean, E. (1997). *Fitting the task to the human: A textbook of Occupational Ergonomics* (5th Ed.). USA: Taylor and Francis.

Krueger. G. P. (1989). Sustained work, fatigue, sleep loss and performance: A review of the issues. *Work and Stress*, 3, 129–141.

Kryter, K. D. (1990). Hearing levels of railroad trainmen. *Laryngoscope*, 100, 1134–1136.

Ku, C-H., & Smith, M. J. (2010). Organisational factors and scheduling in locomotive engineers and conductors: Effects on fatigue, health and social well-being. *Applied Ergonomics*, 41, 62–71.

Kubo, T., Takahashi, M., Liu, X. X., Ikeda, H., Togo, F., Shimazu, A., Tanaka, K., Kamata, N., Kubo, Y., & Uesugi, J. (2016). Fatigue and sleep among employees with prospective increase in work time control: A 1-year observational study with objective assessment. *Journal of Occupational and Environmental Medicine*, 58, 1066–1072. doi: 10.1097/JOM.0000000000000858

Lac, G., & Chamoux, A. (2004). Biological and psychological responses to two rapid shiftwork schedules. *Ergonomics*, 47, 1339–1349. doi: 10.1080/00140130410001724237

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, Article Number: 863. doi: 10.3389/fpsyg.2013.00863

Lambourne, K., & Tomporowski, P. (2010). The effect of exercise-induced arousal on cognitive task performance: A meta-regression analysis. *Brain Research*, 1341, 12–24. doi: 10.1016/j.brainres.2010.03.091

Lamond, N., Dawson, D., & Roach, G. D. (2005). Fatigue assessment in the field: Validation of a hand–held electronic psychomotor vigilance task. *Aviation, Space, and Environmental Medicine*, 76, 486–489.

Lamond, N., Dorrian, J., Roach, G. D., McCulloch, K., Holmes, A. L., Burgess, H. J., Fletcher, A., & Dawson, D. (2003). The impact of a week of simulated night work on sleep, circadian phase, and performance. *Occupational and Environmental Medicine*, 60, 1–9. doi: 10.1136/oem.60.11.e13

Lamond, N., Jay, S. M., Dorrian, J., Ferguson, S. A., Roach, G. D., & Dawson, D. (2008). The sensitivity of a palm–based psychomotor vigilance task to severe sleep loss. *Behavior Research Methods*, 40, 347–352. doi: 10.3758/BRM.40.1.347

Landrigan, C. P., Rothschild, J. M., Cronin, J. W., Kaushal, R., Burdick, E., Katz, J. T., Lilly, C. M., Stone, P. H., Lockley, S. W., Bates, D. W., & Czeisler, C. (2004). Effect of reducing interns' work hours on serious medical errors in intensive care units. NEW *England Journal of Medicine*, 18, 1838-1848. doi: 10.1056/NEJMoa041406

Langner, R., & Eickhoff, S. B. (2013). Sustaining attention to simple tasks: A meta-analytic review of the neural mechanisms of vigilant attention. *Psychological Bulletin*, 139, 870–900. doi: 10.1037/a0030694

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

Lecas, J. C. (2004). Locus coeruleus activation shortens synaptic drive while decreasing spike latency and jitter in sensorimotor cortex. Implications for neuronal integration. *European Journal of Neuroscience*, 19, 2519–2530. doi: 10.1111/j.0953–816X.2004.03341.x

LeCompte, M. D., & Goetz, J. P. (1982). Problems of reliability and validity in ethnographic research. *Review of Educational Research*, 52, 31–60. doi: 10.3102/00346543052001031

Lafrance, C., & Dumont, M. (2000). Diurnal variations in the waking EEG: Comparisons with sleep latencies and subjective alertness. *Journal of Sleep Research*, 9, 243–248.

Lee, S., & Kim, J. K. (2018). Factors contributing to the risk of airline pilot fatigue. *Journal of Air Transport Management*, 67, 197–207. doi: 10.1016/j.jairtraman.2017.12.009

Legault, G., Clement, A., Kenny, G. P., Hardcastle, S., & Keller, N. (2017). Cognitive consequences of sleep deprivation, shiftwork, and heat exposure for underground miners. *Applied Ergonomics*, 58, 144–150. 10.1016/j.apergo.2016.06.007

Lenne, M. G., Triggs, T. J., & Redman, J. R. (1997). Time of day variations in driving performance. *Accident Analysis and Prevention*, 29, 431–437. doi: 10.1016/S0001-4575(97)00022-5

Leproult, R., Colecchia, E. F., Berardi, A. M., Stickgold, R., Kosslyn, S. M., & van Cauter, E. (2003). Individual differences in subjective and objective alertness during sleep deprivation are stable and unrelated. *American Journal of Physiology-Regulatory Integrative and Comparative Physiology*, 284, R280–R290. doi: 10.1152/ajpregu.00197.2002

Leproult, R., Copinschi, G., Buxton, O., & VanCauter, E. (1997). Sleep loss results in an elevation of cortisol levels the next evening. *Sleep*, 20, 865–870.

Leslie, M., Paradis, E., Gropper, M. A., Reeves, S., & Kitto, S. (2014). Applying ethnography to the study of context in healthcare quality and safety. *Bmj Quality & Safety*, 2, 99–105. doi: 10.1136/bmjqs–2013–002335

Leung, M. Y., Bowen, P., Liang, Q., & Famakin, I. (2015). Development of a job-stress model for construction professionals in South Africa and Hong Kong. *Journal of Construction Engineering And Management*, 141, Article number: 04014077. doi: 10.1061/(ASCE)CO.1943-7862.0000934

Levine, T. R., & Hullett, C. R. (2002). Eta Squared, Partial Eta Squared, and misreporting of effect size in communication research. *Human Communication Research,* 28, 612–625. doi: 10.1111/j.1468-2958.2002.tb00828.x

Lewis, T. L., Ellemberg, D., Maurer, D., Dirks, M., Wilkinson, F., & Wilson, H. R. (2004). A window on the normal development of sensitivity to global form in glass patterns. *Perception*, 33, 409–418.

Lewis, G., & Wessely, S. (1992). The epidemiology of fatigue: More questions than answers. *Journal of Epidemiology and Community Health*, 46, 92–97.

Li, Y. Y., Yamamoto, T., & Zhang, G. N. (2018). The effect of fatigue driving on injury severity considering the endogeneity. *Journal of Safety Research*, 64, 11–19. doi: 10.1016/j.jsr.2017.12.007

Lian, Y. L., Gu, Y. Y., Han, R., Jiang, Y., Guan, S. Z., Xiao, J., & Liu, J. W. (2016). Effect of changing work stressors and coping resources on psychological distress. *Journal of Occupational and Environmental Medicine*, 58, 256–263. doi: 10.1097/JOM.0000000000000777

Lie, A., Skogstad, M., Johnsen, T. S., Engdahl, B., & Tambs, K. (2013). Hearing status among Norwegian train drivers and train conductors. *Occupational Medicine–Oxford*, 63, 544–548. doi: 10.1093/occmed/kqt114

Lie, A., Skogstad, M., Johnsen, T. S., Engdahl, B., & Tambs, K. (2016). Noise–induced hearing loss in a longitudinal study of Norwegian railway workers. *BMJ Open*, 6, Article number: e011923. doi: 10.1136/bmjopen–2016–011923

Lim, J., Wu, W. C., Wang, J. J., Detre, J. A., Dinges, D. F., & Rao, H. Y. (2010). Imaging brain fatigue from sustained mental workload: An ASL perfusion study of the time-on-task effect. *Neuroimage*, 49, 3426–3435. doi: 10.1016/j.neuroimage.2009.11.020

Lipscomb, D. M., & Roettger, R. W. (1976). Noise and public health. *Public Health Review*. 5, 67–96.

Littman, A. J., White, E., Satia, J. A., Bowen, D. J., & Kristal, A. R. (2006). Reliability and validity of 2 single-item measures of psychosocial stress. *Epidemiology*, 17, 398–403. doi: 10.1097/01.ede.0000219721.89552.51

Liu, K., Li, B., Qian, S. W., Jiang, Q. J., Li, L., Wang, W., Zhang, G. Y., Sun, Y. J., Sun, G. (2016). Mental fatigue after mild traumatic brain injury: a 3D-ASL perfusion study. *Brain Imaging and Behavior*, 10, 857–868. doi: 10.1007/s11682-015-9492-3

Liu, Q. Y., Shi, J., Duan, P., Liu, B., Li, T. F., Wang, C., Li, H., Yang, T. T., Gan, Y., Wang, X. J., Cao, S. Y., & Lu, Z. X. (2018). Is shift work associated with a higher risk of overweight or obesity? A systematic review of observational studies with meta-analysis. *International Journal of Epidemiology*, 47, 1956–1971. doi: 10.1093/ije/dyy079

Lo, J. C., Groeger, J. A., Santhi, N., Arbon, E. L., Lazar, A. S., Hasan, S., von Schantz, M., Archer, S. N., & Dijk, D. J. (2012). Effects of partial and acute total sleep deprivation on performance across cognitive domains, individuals and circadian phase. *PLoS One*, 7, e45987. doi: 10.1371/journal.pone.0045987

Loh, S., Lamond, N., Dorrian, J., Roach, G., & Dawson, D. (2004). The validity of psychomotor vigilance tasks of less than 10–minute duration. *Behavior Research Methods, Instruments, & Computers*, 36, 339–346. doi: 10.3758/BF03195580

Longo, L. (2015). Designing medical interactive systems via assessment of human mental workload. In: *International Symposium on Computer-Based Medical Systems*. dio: 10.1109/CBMS.2015.67

Longo, L. (2017). Subjective Usability, Mental Workload Assessments and Their Impact on Objective Human Performance. In: *IFIP Conference on Human-Computer Interaction*. dio: 10.1007/978-3-319-67684-5_13

Loudoun, R., & Allan, C. (2008). The effect of time of day on injury patterns amongst adolescents in Australia. *Applied Ergonomics*, 39, 663–670. doi: 10.1016/j.apergo.2007.12.005

Lützhöft, M., Dahlgren, A., Kircher, A., Thorslund, B., & Gillberg, M. (2010). Fatigue at sea in Swedish shipping-a field study. *American Journal of Industrial Medicine*, 53, 733–740.

Ma, L. L., Zhang, Z. L., & Li, M. (2016). Surveying traffic congestion based on the concept of community structure of complex networks. *International Journal of Modern Physics C*, 27, Article number: 1650149. doi: 10.1142/S0129183116501497

Mackworth, J. F. (1968). Vigilance arousal and habituation. *Psychological Review*, 75, 308–322. doi: 10.1037/h0025896

Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology*, 1, 6–21. doi: 10.1080/17470214808416738

Madigan, R., Golightly, D., & Madders, R. (2016). Application of human factors analysis and classification system (HFACS) to UK rail safety of the line incidents. *Accident Analysis and Prevention*, 97, 122–131. doi: 10.1016/j.aap.2016.08.023

Malleson, S. M. (1989). Confirmation of noise exposure levels by analysis of the audiometic data in long service underground train drivers. *Journal of the Society of Occupational Medicine*, 39, 107–110.

Mallis, M. M., Mejdal, S., Nguyen, T. T., & Dinges, D. F. (2004). Summary of the key features of seven biomathematical models of human fatigue and performance. *Aviation Space and Environmental Medicine*, 75, A4–A14.

Mantler, J., Matejicek, A., Matheson, K., & Anisman, H. (2005). Coping with employment uncertainty: A comparison of employed and unemployed workers. *Journal of Occupational Health Psychology*, 10, 200–209.

Martin, J. S., Hebert, M., Ledoux, E., Gaudreault, M., & Laberge, L. (2012). Relationship of chronotype to sleep, light exposure, and work-related fatigue in student workers. *Chronobiology International*, 29, 295–304. doi: 10.3109/07420528.2011.653656

Mason, K. L., Retzer, K. D., Hill, R., & Lincoln, J. M. (2015). Occupational fatalities during the oil and gas boom – United States, 2003-2013. *MMWR–Morbidity and Mortality Weekly Report*, 64, 551–554.

Matheson, A., O'Brien, L., & Reid, J. A. (2014). The impact of shiftwork on health: a literature review. *Journal of Clinical Nursing*, 23, 3309–3320. doi: 10.1111/jocn.12524

Mathisen, G. E., & Bergh, L. I. V. (2016). Action errors and rule violations at offshore oil rigs: The role of engagement, emotional exhaustion and health complaints. *Safety Science*, 85, 130–138. doi: 10.1016/j.ssci.2016.01.008

McCorquodale, B. (2002). *Driver's Reminder Appliance: Effectiveness study: Final report*. London: Railway Safety.

MacDonald, W. (2003). The impact of job demands and workload on stress and fatigue. *Australian Psychologist*, 38, 102–117. doi: 10.1080/00050060310001707107

MacDonald, W., & Bendak, S. (2000). Effects of workload level and 8-versus 12-h workday duration on test battery performance. *International Journal of Industrial Ergonomics*, 26, 399–416. doi: 10.1016/S0169-8141(00)00015-9

McGuffog, A., Spencer, M. B., Stone, B. M., & Turner, C. (2005). *Human factors study of fatigue and shift work. Guidelines for the management and reduction of fatigue in train drivers (T059 Report)*. London: Rail Safety and Standards Board Limited.

McGuffog, A., Spencer, M. B., Turner, C., & Stone, B. M. (2004). *Human factors study of fatigue and shift work. Appendix 1: Working patterns of train drivers: implications for fatigue and safety (T059 Report*). London: Rail Safety and Standards Board Limited.

McIntire, L. K., McKinley, R. A., Nelson, J. M., & Goodyear, C. (2017). Transcranial direct current stimulation versus caffeine as a fatigue countermeasure. *Brain Stimulation*, 10, 1070–1078. doi: 10.1016/j.brs.2017.08.005

McLellan, T. M., Caldwell, J. A., & Lieberman, H. R. (2016). A review of caffeine's effects on cognitive, physical and occupational performance. *Neuroscience and Biobehavioral Reviews*, 71, 294–312. doi: 10.1016/j.neubiorev.2016.09.001

McOrmond, T. (2004). *Changes in working trends over the past decade*. National Statistics, 25–35.

Mehta, R. K., Peres, S. C., Kannan, P., Rhee, J., Shortz, A. E., & Mannan, M. S. (2017). Comparison of objective and subjective operator fatigue assessment methods in offshore shiftwork. *Journal of Loss Prevention in the Process Industries*, 48, 376–381. doi: 10.1016/j.jlp.2017.02.009

Meijman, T. F., Mulder, G., & van Dormolen, M. (1992). Workload of driving examiners: A psychophysiological field study. In H. Kragt (Ed.), *Enhancing industrial performances* (pp. 245–260). London: Taylor & Francis.

Melamed, S., & Bruhis, S. (1996). The effects of chronic industrial noise exposure on urinary cortisol, fatigue and irritability – A controlled field experiment. *Journal of Occupational and Environmental Medicine*. 38, 252–256. doi: 10.1097/00043764–199603000–00009

Mette, J., Garrido, M. V., Harth, V., Preisser, A. M., & Mache, S. (2018). Healthy offshore workforce? A qualitative study on offshore wind employees' occupational strain, health, and coping. *BMC Public Health*, 18, Article number: 172. doi: 10.1186/s12889-018-5079-4

Mitchell, I. (2018). Driver advisory systems – opportunities and challenges. *IRSE News*, 242, 1–4.

Mitchell, J. T., & Dyregrov, A. (1993). Traumatic stress in disaster workers and emergency personnel: Prevention and intervention. In J. P. Wilson., & B. Raphael. (Eds.). *International handbook of traumatic stress syndromes* (pp. 905–914). New York: Plenum Press.

Miyatake, M., & Ko, H. (2010). Optimization of train speed profile for minimum energy consumption. *IEEJ Transactions on Electrical and Electronic Engineering*, 5, 263–269. doi: 10.1002/tee.20528

Monahan, T., & Fisher, J. A. (2010). Benefits of 'observer effects': lessons from the field. *Qualitative Research*, 3, 357–376. doi: 10.1177/1468794110362874

Monfort, S. S., Graybeal, J. J., Harwood, A. E., McKnight, P. E., & Shaw, T. H. (2018). A single-item assessment for remaining mental resources: development and validation of the Gas Tank Questionnaire (GTQ). *Theoretical Issues in Ergonomics Science*, 19, 530–552. doi: 10.1080/1463922X.2017.1397228

Moor, A. (2013). *Category A SPAD and TPWS activity report*. London: Rail Safety and Standards Board Limited.

Moore, L. J., Wilson, M. R., McGrath, J. S., Waine, E., Masters, R. S. W., & Vine, S. J. (2015). Surgeons' display reduced mental effort and workload while performing robotically assisted surgical tasks, when compared to conventional laparoscopy. *Surgical Endoscopy and other Interventional Techniques*, 29, 2553–2560. doi: 10.1007/s00464-014-3967-y

Moreno-Martinez, F. J., & Montoro, P. R. (2012). An ecological alternative to snodgrass & vanderwart: 360 high quality colour images with norms for seven psycholinguistic variables. *PLoS One*, 7, Article number: e37527. doi: 10.1371/journal.pone.0037527

Morris, M. B., Wiedbusch, M. D., & Gunzelmann, G. (2018). Fatigue incident antecedents, consequences, and aviation operational risk management resources. *Aerospace Medicine and Human Performance*, 89, 708–716. doi: 10.3357/AMHP.5019.2018

Motamedzade, M., Abbasinia, M., Parvari, R., Oliaie, M., Karimi, S., & Mohammadi, P. (2017). Mental workload and its association with fatigue in operating room personnel of Hamadan hospitals, Iran, 2016. *Journal of Occupational Health and Epidemiology*, 6, 98–105. doi: 10.29252/johe.6.2.98

Muecke, S. (2005). Effects of rotating night shifts: Literature review. *Journal of Advanced Nursing*, 50, 433–439. doi: 10.1111/j.1365-2648.2005.03409.x

Myrtek, M., Deutschmannjanicke, E., Strohmaier, H., Zimmermann, W., Lawerenz, S., Brugner, G., & Muller, W.: (1994). Physical, mental, emotional, and subjective workload components in train drivers. *Ergonomics*, 37, 1195–1203. doi: 10.1080/00140139408964897

Naghiyev, A., Sharples, S., Ryan, B., Coplestone, A., & Carey, M. (2017). Expert knowledge elicitation to generate human factors guidance for future European Rail Traffic Management System (ERTMS) train driving models. *Proceedings of The Institution of Mechanical Engineers Part F-Journal of Rail and Rapid Transit*, 231, 1141–1149. doi: 10.1177/0954409717695902

Nagy, M. S. (2002). Using a single-item approach to measure facet job satisfaction. *Journal of Occupational and Organizational Psychology*, 75, 77–86. doi: 10.1348/096317902167658

National Highway Traffic Safety Administration. (2010). Assessing the feasibility of vehicle-based sensors to detect alcohol impairment. U.S. Department of Transportation website: www.nhtsa.gov/DOT/NHTSA/NVS/Crash%20Avoidance/.../811358.pdf.

National Institute for Health and Care Excellence (NICE). (2007). Chronic fatigue syndrome/myalgic encephalomyelitis (or encephalopathy): diagnosis and management CG53. National Institute for Health and Care Excellence website: https://www.nice.org.uk

National Institute for Health and Care Excellence (NICE). (2018a). Chronic fatigue syndrome myalgic encephalomyelitis overview. National Institute for Health and Care Excellence website: https://www.nice.org.uk

National Institute for Health and Care Excellence (NICE). (2018b). Assessment and diagnosis of chronic fatigue syndrome myalgic encephalomyelitis. National Institute for Health and Care Excellence website: https://www.nice.org.uk

National Institute for Health and Care Excellence (NICE). (2018c). Management of chronic fatigue syndrome myalgic encephalomyelitis. National Institute for Health and Care Excellence website: https://www.nice.org.uk

National Institute for Health and Care Excellence (NICE). (2018d). Specialist care for chronic fatigue syndrome myalgic encephalomyelitis. National Institute for Health and Care Excellence website: https://www.nice.org.uk

Nawrot, P., Jordan, S., Eastwood, J., Rotstein, J., Hugenholtz, A., & Feeley, M. (2003). Effects of caffeine on human health. *Food Additives and Contaminants Part A-Chemistry Analysis Control Exposure & Risk Assessment*, 20, 1–30. doi: 10.1080/0265203021000007840

Neubauer, A. B., Smyth, J. M., & Sliwinski, M. J. (2018). When you see it coming: Stressor anticipation modulates stress effects on negative affect. *Emotion*, 18, 342–354. doi: 10.1037/emo0000381

Nielsen, B., Hales, J. R. S., Strange, S., Christensen, N. J., Warberg, J., & Saltin, B. (1993): Human circulatory and thermoregulatory adaptations with heat acclimation and exercise in a hot dry environment. *Journal of Physiology–London*, 460, 467–485. doi: 10.1113/jphysiol.1993.sp019482

Nielsen, B., Strange, S., Christensen, N. J., Warberg, J., & Saltin, B. (1997): Acute and adaptive responses in humans to exercise in a warm, humid environment. *Pflügers Archiv–European Journal of Physiology*, 434, 49–56. doi: 10.1007/s004240050361

Neiworth, J. J., Gleichman, A. J., Olinick, A. S., & Lamp, K. E. (2006). Global and local processing in adult humans (Homo sapiens), 5–year–old children (Homo sapiens), and adult cotton–top tamarins (Saguinus oedipus). *Journal of Comparative Psychology*, 120, 323–330.

Nehlig, A. (2010). Is caffeine a cognitive enhancer? *Journal of Alzheimer's Disease*, 20, S85–S94. doi: 10.3233/JAD-2010-091315.

Network Rail. (2012). GSM–R DSD Alarms During Coupling/Uncoupling Activities. Version 1.0. *GSM–R Bulletin*. Retrieved from https://cdn.networkrail.co.uk/wp–content/uploads/2016/12/GSM–R–Bulletin–16–DSD–Alarms–During–Coupling–Uncoupling–Activities–Version–1–new–format.pdf

Network Rail. (2013). Urgent (Yellow Button) Calls to the Signaller. Version 1.0. *GSM–R Bulletin*. Retrieved from https://cdn.networkrail.co.uk/wp–content/uploads/2016/12/GSM–R–Bulletin–24–Urgent–Yellow–Button–Calls–to–the–Signaller–Version–1–new–format.pdf

Network Rail. (2017). ATO – GB New Trains Onboard Subsystem Requirements Specification, STE/ATO/REQ/001 Issue 1.0. Retrieved from https://www.rssb.co.uk/library/improving–industry–performance/ato%20–%20gb–new–train–on–board–subsystems–requirements–specification–ste_ato_req_001–issue–1.pdf

Network Rail. (2018, September 13). Guide to the GSM–R System. *Network Rail*. Retrieved from https://cdn.networkrail.co.uk/wp–content/uploads/2017/06/Network–Rail–Telecoms–briefing–pack.pdf

Neville, K. J., Bisson, R. U., French, J., Boll, P. A., & Storm, W. F. (1994). Subjective fatigue of c-141 aircrews during operation-desert-storm. *Human Factors*, 36, 339–349.

Newington, L., & Metcalfe, A. (2014). Factors influencing recruitment to research: qualitative study of the experiences and perceptions of research teams. BMC *Medical Research Methodology*, 14, Article number: 10. doi: 10.1186/1471-2288-14-10

Nilsson, R., Garling, T., & Lutzhoft, M. (2009). An experimental simulation study of advanced decision support system for ship navigation. *Transportation Research Part F-Traffic Psychology and Behaviour*, 12, 188–197. doi: 10.1016/j.trf.2008.12.005

Nguyen, T. A., & Zeng, Y. (2017). Effects of stress and effort on self-rated reports in experimental study of design activities. *Journal of Intelligent Manufacturing*, 28, 1609–1622. doi: 10.1007/s10845-016-1196-z

Novak, R. D., & Auvil-Novak, S. E. (1996). Focus group evaluation of night nurse shiftwork difficulties and coping strategies. *Chronobiology International*, 13, 457–463. doi: 10.3109/07420529609020916

Noy, Y. I., Horrey, W. J., Popkin, S. M., Folkard, S., Howarth, H. D., & Courtney, T. K. (2011). Future directions in fatigue and safety research. *Accident Analysis and Prevention*, 43, 495–497. doi: 10.1016/j.aap.2009.12.017

Ogino, Y., Kakeda, T., Nakamura, K., & Saito, S. (2014). Dehydration enhances pain–evoked activation in the human brain compared with rehydration. *Anesthesia and Analgesia*, 188, 1317–1325. doi: 10.1213/ANE.0b013e3182a9b028

Öhrström, E., & Björkman, M. (1988). Effects of noise–disturbed sleep – A laboratory study on habituation and subjective noise sensitivity. *Journal of Sound and Vibration*, 122, 277–290.

Öhrström, E., Björkman, M., & Rylander, R. (1990). Effects of noise during sleep with reference to noise sensitivity and habituation. *Environment International*, 16, 477–482.

Ollo-Lopez, A., Bayo-Moriones, A., & Larraza-Kintana, M. (2014). Not All Alike: The Distinctive Impact of Voluntary and Involuntary Effort on Stress and Fatigue. *Applied Research in Quality of Life*, 9, 559–573. doi: 10.1007/s11482-013-9252-8

O'Neil, W. A. (2003). The Human Element in Shipping. *World Maritime University Journal of Maritime Affairs*, 2, 95–97.

Orlandi, L., & Brooks, B. (2018). Measuring mental workload and physiological reactions in marine pilots: Building bridges towards redlines of performance. *Applied Ergonomics*, 69, 74–92. doi: 10.1016/j.apergo.2018.01.005

Office of Rail and Road (ORR). (2012). Managing Rail Staff Fatigue. Office or Rail and Road website: https://orr.gov.uk

Office of Rail and Road (ORR). (2016a). *Points from RSSB Project T1083 regarding the Fatigue & Risk Index*. London: Open Government Licence.

Office of Rail and Road (ORR). (2016b). *Crossrail Programme: Crossrail train protection (Plan B) – Railway Safety Regulations 1999 exemption application report*. Office of Rail and Road website: https://orr.gov.uk/__data/assets/pdf_file/0010/18856/paddington–0–12–exemption–application–report.pdf

Office of Rail and Road (ORR). (2016c). Good practice guidelines – Fatigue factors. Office or Rail and Road website: https://orr.gov.uk

Office of Rail and Road (ORR). (2017a). Business Plan 2017–18. Office of Rail and Road website: https://orr.gov.uk

Office of Rail and Road (ORR). (2017b). Fatigue – Key performance indicators. Office or Rail and Road website: https://orr.gov.uk

Office of Rail and Road (ORR). (ORR). (2018). *Office of Rail and Road: Independent inquiry into the timetable disruption in May 2018 Final Report*. Office of Rail and Road website: https://orr.gov.uk/__data/assets/pdf_file/0010/39916/inquiry–into–may–2018–timetable–disruption–december–2018–report.pdf

Paley, M.J., & Tepas, D.I., (1994). Fatigue and the shiftworker: Firefighters working on a rotating shift schedule. *Human Factors*, 36, 269–284.

Parasuraman, R., de Visser, E., Clarke, E., McGarry, W. R., Hussey, E., Shaw, T., & Thompson, J. C. (2009). Detecting threat-related intentional actions of others: effects of image quality, response mode, and target cuing on vigilance. *Journal of Experimental Psychology-Applied*, 15, 275–290. doi: 10.1037/a0017132

Parkes, K. R. (1995). The effects of objective workload on cognitive performance in a field setting: a two–period cross–over trial. *Applied Cognitive Psychology*. 9, 153–171. doi: 10.1002/acp.2350090710

Parsons, W. D., & Neims, A. H. (1978). Effect of smoking on caffeine clearance. CLINICAL *Pharmacology & Therapeutics*, 24, 40–45.

Patkai, P. (1971). Interindividual differences in diurnal variations in alertness, performance, and adrenaline excretion. *Acta Physiologica Scandinavica*, 81, 35–46. doi: 10.1111/j.1748-1716.1971.tb04875.x

Patrick, G. T. W & Gilbert, J. A. (1896). Studies from the psychological laboratory of the University of IOWA on the effects of loss of sleep. *The Psychological Review*, III, 469–483.

Patterson, P. D., Weaver, M. D., Hostler, D., Guyette, F. X., Callaway, C. W., & Yealy, D. M. (2012). The shift length, fatigue, and safety conundrum in ems. *Prehospital Emergency Care*, 16, 572–576. doi: 10.3109/10903127.2012.704491

Patterson, P. D., Weaver, M. D., Markosyan, M. A., Moore, C. G., Guyette, F. X., Doman, J. M., Sequeira, D. J., Werman, H. A., Swanson, D., Hostler, D., Lynch, J., Templin, M. A., Rozario, N. L., Russo, L., Hines, L., Swecker, K., Runyon, M. S., & Buysse, D. J. (2019). Impact of shift duration on alertness among air-medical emergency care clinician shift workers. *American Journal of Industrial Medicine*, 62, 325–336. doi: 10.1002/ajim.22956

Paxion, J., Galy, E., & Berthelon, C. (2014). Mental workload and driving. *Frontiers in Psychology*, 5, 1344. doi: 10.3389/fpsyg.2014.01344

Pellegrino, P., & Marqueze, E. C. (2019). Aspects of work and sleep associated with work ability in regular aviation pilots. *Revista de Saude Publica*, 53, Article number: 16. doi: 10.11606/S1518-8787.2019053000345

Peter, R., Alfredsson, L., Knutsson, A., Siegrist, J., & Westerholm, P. (1999). Does a stressful psychosocial work environment mediate the effects of shift work on cardiovascular risk factors? *Scandinavian Journal of Work Environment & Health*, 25, 376–381. doi: 10.5271/sjweh.448

Philip, P., & Akerstedt, T. (2006). Transport and industrial safety, how are they affected by sleepiness and sleep restriction. *Sleep Medicine Reviews*, 10, 347–356. doi: 10.1016/j.smrv.2006.04.002

Pink, S., & Morgan, J. (2013). Short–Term Ethnography: Intense Routes to Knowing. *Symbolic Interaction*, 36, 351–361. doi: 10.1002/symb.66

Pilcher, J. J., Nadler, E., & Busch, C. (2002), Effects of hot and cold temperature exposure on performance: A meta–analytic review. *Ergonomics*, 45, 682–698.

Pisarski, A., Lawrence, S. A., Bohle, P., & Brook, C. (2008). Organizational influences on the work life conflict and health of shiftworkers. *Applied Ergonomics*, 39, 580–588.

Portney, L. G., & Watkins, M. P. (2000). *Foundations of clinical research: Applications to practice*. New Jersey: Prentice Hall.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, 13, 25–42. doi: 10.1146/annurev.neuro.13.1.25

Presser, H. B. (2003). Race-ethnic and gender differences in nonstandard work shifts. *Work and Occupations*, 30, 412–439.

Prytz, E. G., & Scerbo, M. W. (2015). Changes in stress and subjective workload over time following a workload transition. *Theoretical Issues in Ergonomics Science*, 16, 586–605. doi: 10.1080/1463922X.2015.1084397

Qian, S. W., Li, M., Li, G. Y., Liu, K., Li, B., Jiang, Q. J., Li, L., Yang, Z., & Sun, G. (2015). Environmental heat stress enhances mental fatigue during sustained attention task performing: Evidence from an ASL perfusion study. *Behavioural Brain Research*, 280, 6–15. doi: 10.1016/j.bbr.2014.11.036

Querstret, D., Cropley, M., Kruger, P., & Heron, R. (2016). Assessing the effect of a Cognitive Behaviour Therapy (CBT)-based workshop on work-related rumination, fatigue, and sleep. *European Journal of Work and Organizational Psychology*, 25, 50–67. doi: 10.1080/1359432X.2015.1015516

Rail Safety and Standards Board (RSSB). (2003). Coping with shift work & fatigue: A good practice guide for drivers. Retrieved from https://www.rssb.co.uk/Library/improving-industry-performance/2003-good-practice-good-shift-and-fatigue.pdf

Rail Safety and Standards Board (RSSB). (2009). *Driver advisory information for energy management and regulation Stage 1 Report*. London: Rail Safety and Standards Board Limited.

Rail Safety and Standards Board (RSSB). (2010a). *Research Programme operations and management: Causes, risks and mitigation of station stopping incidents. Project summary report*. London: Rail Safety and Standards Board Limited.

Rail Safety and Standards Board (RSSB). (2010b). *Driver advisory information for energy management and regulation Stage 2 Report*. London: Rail Safety and Standards Board Limited.

Rail Safety and Standards Board (RSSB). (2012). Managing fatigue: A good practice guide. Retrieved from Rail Safety and Standards Board: http://www.rssb.co.uk/rgs/standards/RS504%20Iss%201.pdf.

Rail Safety and Standards Board (RSSB). (2014a). *Rule book: Module TS7. Signaller token regulations. GE/RT8000/TS7 Rule: Issue 5*. London: Rail Safety and Standards Board Limited.

Rail Safety and Standards Board (RSSB). (2014b*). Requirements for data recorders on trains: Issue 2*. London: Rail Safety and Standards Board Limited.

Rail Safety and Standards Board (RSSB). (2015a). *AWS and TPWS handbook: RS/522 Issue 3*. London: Rail Safety and Standards Board Limited.

Rail Safety and Standards Board (RSSB). (2015b). *Rule book: Train driver manual*. London: Rail Safety and Standards Board Limited.

Rail Safety and Standards Board (RSSB). (2015c). *Guidance on low adhesion between the wheel and the rail – Managing the risk*. London: Rail Safety and Standards Board Limited.

Rail Safety and Standards Board (RSSB). (2015d). *GSM–R User Procedures (Cab Radio). Procedures for using the Siemens GSM–R cab radio: Issue 7.1*. London: Rail Safety and Standards Board Limited.

Rail Safety and Standards Board (RSSB). (2015e). *Rule Book: Module TW1. General safety responsibilities and personal track safety for non–track workers*. London: Rail Safety and Standards Board Limited.

Rail Safety and Standards Board (RSSB). (2015f). *RIS-3751-TOM: Rail industry standard for train driver selection*. London: Rail Safety and Standards Board Limited.

Rail Safety and Standards Board (RSSB). (2016a). *Preparation and movement of trains Issue 11*. London: Rail Safety and Standards Board Limited.

Rail Safety and Standards Board (RSSB). (2016b). *ATOC Guidance Note– Investigation of Station Stopping Incidents ATOC/GN009 Issue 3*. London: Rail Safety and Standards Board Limited.

Rail Safety and Standards Board (RSSB). (2017a). *Rail industry standard RIS–0794–CCS Issue 2. GSM–R train voice radio systems*. London: Rail Safety and Standards Board Limited.

Rail Safety and Standards Board (RSSB). (2017b). *Rule book: GERT8000–Module TW1 Issue 12. Preparation and movement of trains*. London: Rail Safety and Standards Board Limited.

Rail Safety and Standards Board (RSSB). (2018a, November 10). Signals passed at danger (SPADs): A summary of the rail industry's approach to risk reduction. *Rail Safety and Standards Board*. Retrieved from https://www.rssb.co.uk/Library/improving–industry–performance/signals–passed–at–danger–spads–a–summary–of–the–rail–industrys–approach–to–risk–reduction.pdf

Rail Safety and Standards Board (RSSB). (2018b). *Rule book: Train driver manual*. London: Rail Safety and Standards Board Limited.

Rail Safety and Standards Board (RSSB). (2018c, August 12). *Global System for Mobile Communications – Railways (GSM–R). Rail Safety and Standards Board*. Retrieved from https://www.rssb.co.uk/improving–industry–performance/gsm–r

Rail Safety and Standards Board (RSSB). (2018d). *Rule book: GERT8000–Module TW1 Issue 13*. *Preparation and movement of trains*. London: Rail Safety and Standards Board Limited.

Rail Safety and Standards Board (RSSB). (2018e). *Rule book: GERM8000 train driver Issue 6. Train driver manual*. London: Rail Safety and Standards Board Limited.

Rajaratnam, S. M., & Arendt, J. (2001). Health in a 24-h society. *The Lancet*, 358, 999–1005. doi: 10.1016/S0140-6736(01)06108-6

Rajaratnam, S. M. W., & Jones, C. B. (2004). Lessons about sleepiness and driving from the Selby Rail Disaster Case: R v Gary Neil hart. *Chronobiology International*, 6, 1073-1077. doi: 10.1081/LCBI-200035940

Ramakrishnan, S., Wesensten, N. J., Kamimori, G. H., Moon, J. E., Balkin, T. J., & Reifman, J. (2016). A unified model of performance for predicting the effects of sleep and caffeine. *Sleep*, 39, 1827–1841. doi: 10.5665/sleep.6164

Rangan, S., & van Dongen, H. P. A. (2013). Quantifying fatigue risk in model–based fatigue risk management. *Aviation Space and Environmental Medicine*, 84, 155–157. doi: 10.3357/ASEM.3455.2013

Reeves, S., Kuper, A., & Hodges, B. D. (2008). Qualitative research – Qualitative research methodologies: ethnography. *BMJ–British Medical Journal*, 337, Article Number: a1020. doi: 10.1136/bmj.a1020

Reid, K. J., Weng, J., Ramos, A. R., Zee, P. C., Daviglus, M., Mossavar-Rahmani, Y., Sotres-Alvarez, D., Gallo, L. C., Chirinos, D. A., & Patel, S. R. (2018). Impact of shift work schedules on actigraphy-based measures of sleep in Hispanic workers: results from the Hispanic Community Health Study/Study of Latinos ancillary sueño study. *Sleep*, 41, Article number: UNSP zsy131. doi: 10.1093/sleep/zsy131

Reimers, S., & Stewart, N. (2007). Adobe Flash as a medium for online experimentation: A test of reaction time measurement capabilities. *Behavior Research Methods*, 39, 365–370. dio: 10.3758/BF03193004.

Remmen, L. N., Herttua, K., Riss-Jepsen, J., & Berg-Beckhoff, G. (2017). Fatigue and workload among Danish fishermen. *International Maritime Health*, 68, 252–259. doi: 10.5603/IMH.2017.0044

Richards, J. R., Stayton, T. L., Wells, J. A., Parikh, A. K., & Laurin, E. G. (2018). Night shift preparation, performance, and perception: are there differences between emergency medicine nurses, residents, and faculty? *Clinical and Experimental Emergency Medicine*, 5, 240–248. doi: 10.15441/ceem.17.270

Riethmeister, V., Brouwer, S., van der Klink, J., & Bultmann, U. (2016). Work, eat and sleep: towards a healthy ageing at work program offshore. *BMC Public Health*, 16, Article number: 134. doi: 10.1186/s12889-016-2807-5

Ritson, N., & Charlton, M. (2006). *Health and shiftworking in an administrative environment. Journal of Managerial Psychology*, 21, 131–144.

Rizzo, A. A., Stamps, L. E., & Fehr, La. (1988). Effects of caffeine withdrawal on motor-performance and heart-rate changes. *International Journal of Psychophysiology*, 6, 9–14. doi: 10.1016/0167-8760(88)90030-X

Roach, G. D., Dawson, D., & Lamond, N. (2006). Can a shorter psychomotor vigilance task be used as a reasonable substitute for the ten-minute psychomotor vigilance task? *Chronobiology International*, 23, 379–387. doi: 10.1080/07420520601067931

Roach, G. D., Petrilli, R. M. A., Dawson, D., & Lamond, N. (2012). Impact of layover length on sleep, subjective fatigue levels, and sustained attention of long-haul airline pilots. *Chronobiology International*, 29, 580–586. doi: 10.3109/07420528.2012.675222

Robbins, T. W. (1984). Cortical noradrenaline, attention and arousal. *Psychological Medicine*, 14, 13–21. doi: 10.1017/S0033291700003032

Robertson, D., Wade, D., Workman, R., Woosley, R. L., & Oates, J. A. (1981). Tolerance to the humoral and hemodynamic: effects of caffeine in man. *Journal of Clinical Investigation*, 67, 1111–1117. doi: 10.1172/JCI110124

Robinson, L. J., Gallagher, P., Watsonid, S., Pearce, R., Finkelmeyer, A., Maclachlan, L., & Newton, J. L. (2019). Impairments in cognitive performance in chronic fatigue syndrome are common, not related to co-morbid depression but do associate with autonomic dysfunction. *PLoS One*, 14, e0210394. 10.1371/journal.pone.0210394

Rogers, M., Spencer, M., & Stone, B. (1999). *Validation and development of a method for assessing the risks arising from mental fatigue*. Hampshire, UK: Health & Safety Executive.

Roscoe, A. H. (1978). Stress and workload in pilots. *Aviation Space and Environmental Medicine*, 49, 630–636.

Rosekind, M. R., Gander, P. H., Gregory, K. B., Smith, R. M., Miller, D. L., Oyung, R., Webbon, L. L., & Johnson, J. M. (1996). Managing fatigue in operational settings .2. An integrated approach. *Behavioral Medicine*, 21, 166–170. doi: 10.1080/08964289.1996.9933754

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic–level object recognition. *Perception*, 33, 217–236. doi: 10.1068/p5117

Roth, T. (2012). Appropriate therapeutic selection for patients with shift work disorder. *Sleep Medicine*, 13, 335–341. doi: 10.1016/j.sleep.2011.11.006

Rudari, L., Johnson, M. E., Geske, R. C., & Sperlak, L. A. (2016). Pilot perceptions on impact of crew rest regulations on safety and fatigue. *International Journal of Aviation Aeronautics and Aerospace*, 3, Article number: UNSP 4. doi: 10.15394/ijaaa.2016.1096

Sadeghniiat-Haghighi, K., & Yazdi, Z. (2015). Fatigue management in the workplace. *Industrial Psychiatry Journal*, 24, 12–17. doi: 10.4103/0972-6748.160915

Safrit, M. J. E. (1975). *Reliability theory*. Washington, DC: American Alliance for Health, Physical Education, and Recreation.

Sagaspe, P., Taillard, J., Åkerstedt, T., Bayon, V., Espie, S., Chaumet, G., Bioulac, B., & Philip, P. (2008). Extended driving impairs nocturnal driving performances. *PLoS One*, 3, e3493. doi: 10.1371/journal.pone.0003493

Saito, S. (1992). Does fatigue exist in a quantitative measurement of eye-movements. *Ergonomics*, 35, 607–615 dio: 10.1080/00140139208967840

Salyga, J., & Kusleikaite, M. (2011). Factors influencing psychoemotional strain and fatigue, and relationship of these factors with health complaints at sea among lithuanian seafarers. *Medicina-Lithuania*, 47, 675–681. doi: 10.3390/medicina47120099

Samn, S. W., & Perelli, L. P. (1982). *Estimating aircrew fatigue: A technique with application to airlift operations. Technical Report SAM-TR-82-21*. USAF School of Medicine, Texas, USA: Brooks AFB.

Sanquist, T. F., Raby, M., Forsythe, A., & Carvalhais, A. B. (1997). Work hours, sleep patterns and fatigue among merchant marine personnel. *Journal of Sleep Research*, 6, 245–251.

Santamarina-Perez, P., Eiroa-Orosa, F. J., Rodriguez-Urrutia, A., Qureshi, A., & Alegre, J. (2014). Neuropsychological impairment in female patients with chronic fatigue syndrome: A preliminary study. *Applied Neuropsychology-Adult*, 21, 120–127. doi: 10.1080/09084282.2013.771264

Saper, C. B., Scammell, T. E., & Lu, J. (2005). Hypothalamic regulation of sleep and circadian rhythms. *Nature*, 437, 1257–1263. doi: 10.1038/nature04284

Sasaki, S., Yoshioka, E., Saijo, Y., Bannai, A., Kita, T., Tamakoshi, A., & Kishi, R. (2018). A prospective cohort study of insomnia and chronic kidney disease in Japanese workers. *Sleep and Breathing*, 22, 257–265. doi: 10.1007/s11325-017-1508-3

Satterfield, B. C., & van Dongen, H. P. A. (2013). Occupational fatigue, underlying sleep and circadian mechanisms, and approaches to fatigue risk management. *Fatigue: Biomedicine, Health & Behavior*, 1, 118–136. doi: 10.1080/21641846.2013.798923

Satterfield, B. C., Wisor, J. P., Schmidt, M. A., & van Dongen, H. P. A. (2017). Time-on-task effect during sleep deprivation in healthy young adults is modulated by dopamine transporter genotype. *Sleep*, 40, Article number: UNSP zsx167, doi: 10.1093/sleep/zsx167

Savage, J. (2000a). Ethnography and health care. *British Medical Journal*, 321, 1400–1402. doi: 10.1136/bmj.321.7273.1400

Savage, J. (2000b). Participative observation: Standing in the shoes of others? *Qualitative Health Research*, 3, 324–339. doi: 10.1177/104973200129118471

Savvidis, C., & Koutsilieris, M. (2012). Circadian rhythm disruption in cancer biology. *Molecular Medicine*, 18, 1249–1260. doi: 10.2119/molmed.2012.00077

Schnall, P. L., Belkic, K., Landsbergis, P., & Baker, D. (2000). The workplace and cardiovascular disease. *Occupational Medicine: State Art Review*, 15, 7–68.

Schock, A., Ryan, B., Wilson, J., Clarke, T., & Sharples, S. (2010). Visual scenario analysis: understanding human factors of planning in rail engineering. *Production Planning & Control*, 21, 386–398. doi: 10.1080/09537280903454123

Schriesheim, C. A., Hinkin, T. R., & Podsakoff, P. M. (1991). Can ipsative and single-item measures produce erroneous results in field studies of French and Raven's (1959) five bases of power? An empirical investigation. *Journal of Applied Psychology*, 76, 106–114. doi: 10.1037/0021-9010.76.1.106

Schubert, T. W., Murteira, C., Collins, E. C., & Lopes, D. (2013). ScriptingRT: A software library for collecting response latencies in online studies of cognition. *PLoS One*, 8, Article number: e67769. doi: 10.1371/journal.pone.0067769

Searle, B., Bright, J. E. H., & Bochner, S. (2001). Helping people to sort it out: The role of social support in the job strain model. *Work & Stress*, 15, 328–346. doi: 10.1080/02678370110086768

Seider, A., Wagner, M., Schubert, M., Dröge, P., Römer, K., Pons-Kühnemann, J., Swart, E., Zeeb, H., & Hegewald, J. (2016). Aircraft, road and railway traffic noise as risk factors for heart failure and hypertensive heart disease: A case control study based on secondary data. *International Journal of Hygiene and Environmental Health*, 219, 749–758. doi: 10.1016/j.ijheh.2016.09.012

Seidl, R., Peyrl, A., Nicham, R., & Hauser, E. (2000). A taurine and caffeine-containing drink stimulates cognitive performance and well-being. *Amino Acids*, 19, 635–642. doi: 10.1007/s007260070013

Seo, S. N. (2017). Beyond the Paris Agreement: Climate change policy negotiations and future directions. *Regional Science Policy and Practice*, 9, 121–141. doi: 10.1111/rsp3.12090

Shakouri, M., Ikuma, L. H., Aghazadeh, F., & Nahmens, I. (2018). Analysis of the sensitivity of heart rate variability and subjective workload measures in a driving simulator: The case of highway work zones. *International Journal of Industrial Ergonomics*, 66, 136–145. doi: 10.1016/j.ergon.2018.02.015

Shaw, E. P., Rietschel, J. C., Hendershot, B. D., Pruziner, A. L., Miller, M. W., Hatfield, B. D., & Gentili, R. J. (2018). Measurement of attentional reserve and mental effort for cognitive workload assessment under various task demands during dual-task walking. *Biological Psychology*, 134, 39–51. doi: 10.1016/j.biopsycho.2018.01.009

Shockey, T. M., & Wheaton, A. G. (2017). Short sleep duration by occupation group – 29 states, 2013-2014. *MMWR-Morbidity and Mortality Weekly Report*, 66, 207–213.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations - uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428. doi: 10.1037/0033-2909.86.2.420

Signal, T. L., Mulrlne, H. M., van den Berg, M. J., Smith, A. A. T., Gander, P. H., & Serfontein, W. (2014). Mitigating and monitoring flight crew fatigue on a westward ultra-long-range flight. *Aviation Space and Environmental Medicine*, 85, 1199–1208. doi: 10.3357/ASEM.4034.2014

Simcox, T., & Fiez, J. A. (2014). Collecting response times using Amazon Mechanical Turk and Adobe Flash. *Behavior Research Methods*, 46, 95–111. doi: 10.3758/s13428–013–0345–y

Simpson, N. S., Diolombi, M., Scott-Sutherland, J., Yang, H., Bhatt, V., Gautam, S., Mullington, J., & Haack, M. (2016). Repeating patterns of sleep restriction and recovery: Do we get used to it? *Brain Behavior and Immunity*, 58, 142–151. doi: 10.1016/j.bbi.2016.06.001

Smith, A. P. (1992). Time of day and performance. In: A. P. Smith, & D. M. Jones (Eds.), *Handbook of human performance*. Vol. 3. State and trait (pp. 217–235). London: Academic Press.

Smith, A. (2002). Effects of caffeine on human behavior. *Food and Chemical Toxicology*, 40, 1243–1255. doi: 10.1016/S0278-6915(02)00096-0

Smith, A. P. (2005). Caffeine at work. *Human Psychopharmacology: Clinical and Experimental*, 20: 441–445. doi: 10.1002/hup.705

Smith, A. P. (2008). A diary study of fatigue offshore. In: P. D. Bust (Eds.), *International Conference on Contemporary Ergonomics (CE2008)* (pp. 602–606). Nottingham, UK: CRC Press.

Smith, A. P. (2013). Caffeine, extraversion and working memory. *Journal of Psychopharmacology*, 27, 71–76. doi: 10.1177/0269881112460111

Smith, A. P. (2016). A UK survey of driving behaviour, fatigue, risk taking and road traffic accidents. *BMJ Open*, 6, Article number: e011461. doi: 10.1136/bmjopen–2016–011461

Smith, A. P., & Allen, P. H. (2013). Fatigue in the maritime and road haulage industries. In: M. Anderson (Eds.), *International Conference on Contemporary Ergonomics and Human Factors of the Institute of Ergonomics and Human Factors* (pp. 183–188). Cambridge, England: University of Cambridge.

Smith, A. P., Allen, P. H., & Wadsworth, E. J. K. (2015). Crew, manning and fatigue. In D. Pockett., & D. Patraiko (Ed.), *Navigation accidents and their causes* (pp. 1–7). London: Nautical Institute.

Smith, A. P., Allen, P. H., & Wadsworth, E. J. K. (2006). *Seafarer fatigue: The Cardiff research programme*. Centre for Occupational and Health Psychology, Cardiff University website: http://orca.cf.ac.uk/48167/1/research_report_464.pdf.

Smith, A., Brice, C., Nash, J., Rich, N., & Nutt, D. J. (2003). Caffeine and central noradrenaline: effects on mood, cognitive performance, eye movements and cardiovascular function. *Journal of Psychopharmacology*, 17, 283–292. doi: 10.1177/02698811030173010

Smith, A. P., Borysiewicz, L., Pollock, J., Thomas, M., Perry, K., & Llewelyn, M. (1999). Acute fatigue in chronic fatigue syndrome patients.*Psychological Medicine*, 29, 283–290. doi: 10.1017/S003329179800748X

Smith, A. P., Christopher, G., & Sutherland, D. (2013). Acute effects of caffeine on attention: a comparison of non–consumers and withdrawn consumers. *Journal of Psychopharmacology*, 27, 77–83. doi: 10.1177/0269881112460112

Smith, A., & Lane, T. (2001). Fatigue offshore: A comparison of short sea shipping and the offshore oil industry. In: P. McCabe (Eds.), *Contemporary ergonomics* (pp. 467–472). London: Taylor & Francis.

Smith, A., & Nutt, D. (1996). Noradrenaline and attention lapses. *Nature*, 380, 291–291. doi: 10.1038/380291a0

Smith, A. P., & Smith, H. (2017a). Workload, fatigue and performance in the rail industry. In: L. Longo, & M. C. Leva (Eds.), *Human Mental Workload: Models and Applications. H–WORKLOAD 2017. Communications in Computer and Information Science* (pp. 251–263). Cham: Springer.

Smith, A. P. and Smith, H. (2017b). Effects of noise on the well–being of railway staff. *12th ICBEN Congress on Noise as a Public Health Problem*, article number: 2460.

Smith, A. P., & Smith, H. (2017c). A short questionnaire to measure wellbeing at work (Short-SWELL) and to examine the interaction between the employee and organisation. In: R. Charles, & J. Wilkinson (Eds.), *Contemporary Ergonomics and Human Factors 2017* (pp. 200-205). Leicester: Chartered Institute of Ergonomics and Human Factors.

Smith, A., Sutherland, D., & Christopher, G. (2005). Effects of repeated doses of caffeine on mood and performance of alert and fatigued volunteers. *Journal of Psychopharmacology*, 19, 620–626. doi: 10.1177/0269881105056534

Smith, A., Whitney, H., Thomas, M., Perry, K., & Brockman, P. (1997). Effects of caffeine and noise on mood, performance and cardiovascular functioning. *Human Psychopharmacology - Clinical and Experimental*, 12, 27–33. doi: 10.1002/(SICI)1099-1077(199701/02)12:1<27::AID-HUP827>3.3.CO;2-P

Smith, A. P., Wilson, S. J., Glue, P., & Nutt, D. J. (1992). The effects and after effects of the α2-adrenoceptor antagonist idazoxan on mood, memory and attention in normal volunteers. *Journal of Psychopharmacology*, 6, 376–381. doi: 10.1177/026988119200600306

Smith, B. P., Browne, M., Armstrong, T. A., & Ferguson, S. A. (2016). The accuracy of subjective measures for assessing fatigue related decrements in multi-stressor environments. *Safety Science*, 86, 238–244. doi: 10.1016/j.ssci.2016.03.006

Smolarek, L., & Soliwoda, J. (2014). Human fatigue model at maritime transport. In S. Martorell., C. Guedes Soare., & J. Barnett (Ed.), *Safety, reliability and risk analysis: Theory, methods and applications volume 1* (pp. 3295–3302). London: Taylor & Francis Group.

Snel, J., & Lorist, M. M. (2011). Effects of caffeine on sleep and cognition. *Progress in Brain Research*, 190, 105–117. doi: 10.1016/B978-0-444-53817-8.00006-2

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory*, 6, 174–215.

Snyder, E., Cai, B., DeMuro, C., Morrison, M. F., & Ball, W. (2018). A new single-item sleep quality scale: results of psychometric evaluation in patients with chronic primary insomnia and depression. *Journal of Clinical Sleep Medicine*, 14, 1849–1857. doi: 10.5664/jcsm.7478

So, W. K. Y., Wong, S. W. H., Mak, J. N., & Chan, R. H. M. (2017). An evaluation of mental work-load with frontal EEG. *PLoS One*, 12, e0174949. doi: 10.1371/journal.pone.0174949

Somvang, V., Hayward, B., & Cabon, P. (2016). *Preparing rail industry guidance on biomathematical fatigue models*. London: Rail Safety and Standards Board Limited.

Sparrow, A. R., Mollicone, D. J., Kan, K., Bartels, R., Satterfield, B. C., Riedy, S. M., Unice, A., & van Dongen, J. P. A. (2016). Naturalistic field study of the restart break in US commercial motor vehicle drivers: truck driving, sleep, and fatigue. *Accident Analysis and Prevention*, 93, 55–64. doi: 10.1016/j.aap.2016.04.019

Spencer, M. B., Robertson, K. A., & Folkard, S. (2006). *The development of a fatigue / risk index for shiftworkers*. Norwich, UK: Health & Safety Executive.

Stanton, N. A., Salmon, P., Jenkins, D., & Walker, G. (2010). *Human factors in the design and evaluation of central control room operations*. Florida, USA: Taylor & Francis Group.

Sundstrom, E., Town, J. P., Rice, R. W., Osborn, D. P., & Brill, M. (1994). Office noise, satisfaction, and performance. *Environment and Behavior*. 26, 195–222. doi: 10.1177/001391659402600204

Supriyanto, A., Soemarko, D. S., & Prihartono, J. (2018). Hazard function analysis of shiftwork's effect on the incidence of metabolic syndrome among male workers in a chemical company. *Advanced Science Letters*, 24, 6610–6615. doi: 10.1166/asl.2018.12794

Sussman, D., & Coplen, M. (2000). Fatigue and alertness in the United States railroad industry part I: the nature of the problem. *Transportation Research Part F: Traffic Psychology and Behaviour*, 3, 211–220.

The British Psychological Society (BPS). (2014). *Code of human research ethics*. Leicester: The British Psychological Society.

The Royal Society for the Prevention of Accidents (RoSPA). (2017). Driver fatigue and road accidents factsheet. Road safety factsheet website: https://www.rospa.com/rospaweb/docs/advice-services/road-safety/drivers/driver-fatigue-factsheet.pdf

Thomas, M. A., & Smith, A. P. (2006). An investigation of the long-term benefits of antidepressant medication in the recovery of patients with chronic fatigue syndrome. *Human Psychopharmacology: Clinical and Experimental*, 21, 503–509. doi: 10.1002/hup.805

Thomas, M. A., & Smith, A. P. (2009). An Investigation into the cognitive deficits associated with chronic fatigue syndrome. *The Open Neurology Journal*, 3, 13–23. doi: 10.2174/1874205X00903010013

Thomas, M. J. W., Paterson, J. L., Jay, S. M., Matthews, R. W., & Ferguson, S. A. (2019). More than hours of work: fatigue management during high–intensity maritime operations. *Chronobiology International*, 36, 143–149. doi: 10.1080/07420528.2018.1519571

Toci, A. T., Farah, A., Pezza, H. R., & Pezza, L. (2016). Coffee Adulteration: More than Two Decades of Research. *Critical Reviews in Analytical Chemistry*, 46, 83–92. doi: 10.1080/10408347.2014.966185

Toffoli, A., Lefevre, J. M., Bitner–Gregersen, E., & Monbaliu, J. (2005). Towards the identification of warning criteria: Analysis of a ship accident database. *Applied Ocean Research*, 27, 281–291. doi: 10.1016/j.apor.2006.03.003

Tononi, G., & Cirelli, C. (2006). Steep function and synaptic homeostasis. *Sleep Medicine Reviews*, 10, 49–62. doi: 10.1016/j.smrv.2005.05.002

Tsao, L. X., Chang, J., & Ma, L. (2017). Fatigue of Chinese railway employees and its influential factors: Structural equation modelling. *Applied Ergonomics*, 62, 131–141. doi: 10.1016/j.apergo.2017.02.021

Turner, C., & Stone, B. M. (2004). *Human factors study of fatigue and shift work. Appendix 2: Review of coping strategies to mitigate fatigue of train drivers (T059 Report)*. London: Rail Safety and Standards Board Limited.

Tucker, P. (2003). The impact of rest breaks upon accident risk, fatigue and performance: A review. *Work and Stress*, 17, 123–137. doi: 10.1080/0267837031000155949

Tucker, P., Smith, L., Macdonald, I., & Folkard, S. (1998). Shift length as a determinant of retrospective on-shift alertness. *Scandinavian Journal of Work Environment & Health*, 24, 49–54.

Tzannatos, E. (2010). Human element and accidents in Greek shipping. *Journal of Navigation*, 63, 119–127. doi: 10.1017/S0373463309990312

United Nations Framework Convention on Climate Change (UNFCCC). (2015). *The Paris Agreement*. New York: UNFCCC.

van den Berg, M. J., Signal, T. L., Mulrine, H. M., Smith, A. A. T., Gander, P. H., & Serfontein, W. (2015). Monitoring and managing cabin crew sleep and fatigue during an ultra-long range trip. *Aerospace Medicine and Human Performance*, 86, 705–713. doi: 10.3357/AMHP.4268.2015

van den Hombergh, P., Kuenzi, B., Elwyn, G., van Doremalen, J., Akkermans, R., Grol, R., & Wensing, M. (2009). High workload and job stress are associated with lower practice performance in general practice: an observational study in 239 general practices in the Netherlands. *BMC Health Services Research,* 9, Article number: 118. doi: 10.1186/1472-6963-9-118

van der Meij, L., Gubbels, N., Schaveling, J., Almela, M., & van Vugt, M. (2018). Hair cortisol and work stress: Importance of workload and stress model (JDCS or ERI). *Psychoneuroendocrinology*, 89, 78–85. doi: 10.1016/j.psyneuen.2017.12.020

van der Ploeg, E., & Kleber, R. J. (2003). Acute and chronic job stressors among ambulance personnel: predictors of health symptoms. *Occupational and Environmental Medicine*, 60, 40–46.

van Dongen, H. P. A. (2006). Shift work and inter-individual differences in sleep and sleepiness. *Chronobiology International*, 23, 1139–1147. doi: 10.1080/07420520601100071

van Dongen, H. P., & Dinges, D. F. (2005). Sleep, circadian rhythms, and psychomotor vigilance. *Clinical Journal of Sport Medicine*, 24, 237–249.

van Dongen, H. P. A., Maislin, G., Mullington, J. M., & Dinges, D. F. (2003). The cumulative cost of additional wakefulness: Dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep*, 26, 117–126, doi: 10.1093/sleep/26.2.117

van Hooff, M. L. M., Geurts, S. A. E., Kompier, M. A. J., & Taris, T. W. (2007). "How fatigued do you currently feel?" convergent and discriminant validity of a single-item fatigue measure. *Journal of Occupational Health*, 49, 24–234. doi: 10.1539/joh.49.224

van Kamp, I., & Davies, H. (2013). Noise and health in vulnerable groups: A review. *Noise & Health*, 15, 153–159. doi: 10.4103/1463–1741.112361

van Maanen, M. (1979). The fact of fiction in organizational ethnography. *Administrative Science Quarterly*, 24, 539–550.

van Ryswyk, E., Weeks, R., Bandick, L., O'Keefe, M., Vakulin, A., Catcheside, P., Barger, L., Potter, A., Poulos, N., Wallace, J., & Antic, N. A. (2017). A novel sleep optimisation programme to improve athletes' well-being and performance. *European Journal of Sport Science*, 17, 144–151. doi: 10.1080/17461391.2016.1221470

van Steenbergen, H. (2016, June 16). Retiring the QRTEngine. Retrieved from URL http://www.qrtengine.com

Vergara, R. C., Moenne-Loccoz, C., & Maldonado, P. E. (2017). Cold-Blooded attention: Finger temperature predicts attentional performance. *Frontiers in Human Neuroscience*, 11, 454. doi: 10.3389/fnhum.2017.00454

Verster, J. C., & Roth, T. (2013). Vigilance decrement during the on-the-road driving tests: The importance of time-on-task in psychopharmacological research. *Accident Analysis and Prevention*, 58, 244–248. doi: 10.1016/j.aap.2012.10.005

Volker, I., Kirchner, C., & Bock, O. L. (2016). On the relationship between subjective and objective measures of fatigue. *Ergonomics*, 59, 1259–1263. doi: 10.1080/00140139.2015.1110622

von Bastian, C. C., Locher, A., & Ruflin, M. (2013). Tatool: A Java–based open–source programming framework for psychological studies. *Behavior Research Methods*, 45, 108–115. doi: 10.3758/s13428- 012-0224-y

Wadsworth, E. J. K., Allen, P. H., McNamara, R. L., & Smith, A. P. (2008). Fatigue and health in a seafaring population. *Occupational Medicine*, 58,198–204.

Wadsworth, E. J. K., Allen, P. H., Wellens, B. T., McNamara, R. L., & Smith, A. P. (2006). Patterns of fatigue among seafarers during a tour of duty. *American Journal of Industrial Medicine*, 49, 836–844.

Walasa, W. M., Carey, R. N., Fritschi, L., Heyworth, J. S., Fernandez, R. C., & Boyle, T. (2018). Association between shiftwork and the risk of colorectal cancer in females: a population-based case-control study. *Occupational and Environmental Medicine*, 75, 344–350. doi: 10.1136/oemed-2017-104657

Walsh, J. K., Muehlbach, M. J., & Schweitzer, P. K. (1995). Hypnotics and caffeine as countermeasures for shiftwork-related sleepiness and sleep disturbance. *Journal of Sleep Research*, 4, 80–83. doi: 10.1111/j.1365-2869.1995.tb00233.x

Walsh, M. M., Gunzelmann, G., & van Dongen, H. P. A. (2017). Computational cognitive modeling of the temporal dynamics of fatigue from sleep loss. *Psychonomic Bulletin & Review*, 24, 1785–1807. doi: 10.3758/s13423-017-1243-6

Wang, Y. H., Tang, T., Ning, B., van den Boom, T. J. J., & De Schutter, B. (2015). Passenger–demands–oriented train scheduling for an urban rail transit network. *Transportation Research Part C–Emerging Technologies*, 60, 1–23. doi: 10.1016/j.trc.2015.07.012

Wang, G. Y., Xiao, S., Chen, X., & Li, X. (2018). Application of genetic algorithm in automatic train operation. *Wireless Personal Communications*, 102, 1695–1704. doi: 10.1007/s11277-017-5228-6

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors*, 50, 433-441. doi: 10.1518/001872008X312152

Waterhouse, B. D., Azizi, S. A., Burne, R. A., & Woodward, D. J. (1990). Modulation of rat cortical area–17 neuronal responses to moving visual-stimuli during norepinephrine and serotonin microiontophoresis. *Brain Research*, 514, 276–292. doi: 10.1016/0006–8993(90)91422–D

Waterhouse, B. D., Sessler, F. M., Cheng, J. T., Woodward, D. J., Azizi, S. A., & Moises, H. C. (1988). New evidence for a gating action of norepinephrine in central neuronal circuits of mammalian brain. *Brain Research Bulletin*, 21, 425–432. doi: 10.1016/0361–9230(88)90154–2

Watson, P., Whale, A., Mears, S. A., Reyner, L. A., & Maughan, R. J. (2015). Mild hypohydration increases the frequency of driver errors during a prolonged, monotonous driving task. *Physiology & Behavior*, 147, 313–318. doi: 10.1016/j.physbeh.2015.04.028

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research*, 19, 231–240. doi: 10.1519/15184.1

Widyanti, A., Hanna., Muslim, K., & Sutalaksana, I. Z. (2017). The sensitivity of Galvanic Skin Response for assessing mental workload in Indonesia. Work: A *Journal of Prevention Assessment & Rehabilitation*, 56, 111–117. doi: 10.3233/WOR-162479

Wilkinson, R. T., & Houghton, D. (1975). Portable four-choice reaction time test with magnetic tape memory. *Behavior Research Methods & Instrumentation*, 7, 441–446. doi: 10.3758/BF03201557

Williams, G. M., & Smith, A. P. (2018). A practical method of predicting wellbeing at work: The Wellbeing Process Tool. *Advances in Social Sciences Research Journal*, 5, 86–92. doi: 10.14738/assrj.52.4158

Williamson, A. M., Feyer, A. M., & Friswell, R. (1996). The impact of work practices on fatigue in long distance truck drivers. *Accident Analysis & Prevention*, 28, 709–719.

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. doi: 10.1016/j.aap.2009.11.011

Winek, C. L., & Esposito, F. M. (1985). Blood alcohol concentrations: Factors affecting predictions. *Legal Medicine*, 34–61.

Wittmann, M., Dinich, J., Merrow, M., & Roenneberg, T. (2006). Social jetlag: Misalignment of biological and social time. *Chronobiology International*, 23, 497–509. doi: 10.1080/07420520500545979

Wong, J. Y., Lin, J. H., Liu, S. H., & Wan, T. H. (2014). Fireman's job stress: Integrating work/non-work conflict with job demand-control-support model. *European Review of Applied Psychology-Revue Europeenne De Psychologie Appliquee*, 64, 83–91. doi: 10.1016/j.erap.2013.12.002

Woods, D. L., Wyma, J. M., Yund, E. W., Herron, T. J., & Reed, B. (2015). Age-related slowing of response selection and production in a visual choice reaction time task. *Frontiers in Human Neuroscience*, 9, Article Number: 193. doi: 10.3389/fnhum.2015.00193

Woods, M. (2012). *Research into station design and crowd management (T916)*. London: Rail Safety and Standards Board Limited.

Woods, M., & Barrett, A. (2018). *Past research into the health and safety effects of crowding*. London: Rail Safety and Standards Board Limited.

World Health Organization (WHO). (2015). Global status report on road safety 2015. World Health Organization website: https://www.who.int/

World Medical Association. (2013). World Medical Association Declaration of Helsinki Ethical Principles for Medical Research Involving Human Subjects. *Jama–Journal of the American Medical Association*, 310, 2191–2194.

Wyatt, J. K., Cajochen, C., Ritz-De Cecco, A., Czeisler, C. A., & Dijk, D. J. (2004). Low-dose repeated caffeine administration for circadian-phase-dependent performance degradation during extended wakefulness. *Sleep*, 27, 374–381. doi: 10.1093/sleep/27.3.374

Xu, J. B., Potter, M., Tomas, C., Elson, J. L., Morten, K. J., Poulton, J., Wang, N., Jin, H. Q., Hou, Z. X., & Huang, W. E. (2019). A new approach to find biomarkers in chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME) by single-cell Raman micro-spectroscopy. *Analyst*, 144, 913–920. doi: 10.1039/c8an01437

Xu, J. L., Min, J. L., & Hu, J. F. (2018). Real-time eye tracking for the assessment of driver fatigue. *Healthcare Technology Letters*, 5, 54–58. dio: 10.1049/htl.2017.0020

Yen, J. R., Hsu, C. C., Yang, H., & Ho, H. (2009). An investigation of fatigue issues on different flight operations. *Journal of Air Transport Management*, 15, 236–240. doi: 10.1016/j.jairtraman.2009.01.001

Yin, J. T., Tang, T., Yang, L. X., Xun, J., Huang, Y. R., & Gao, Z. Y. (2017). Research and development of automatic train operation for railway transportation systems: A survey. *Transportation Research Part C–Emerging Technologies*, 85, 548–572. doi: 10.1016/j.trc.2017.09.009

Yong, L. C., Li, J., & Calvert, G. M. (2017). Sleep-related problems in the US working population: Prevalence and association with shiftwork status. *Occupational and Environmental Medicine*, 74, 93–104. doi: 10.1136/oemed-2016-103638

Zammouri, A., Moussa, A. A., & Mebrouk, Y. (2018). Brain-computer interface for workload estimation: Assessment of mental efforts in learning processes. *Expert Systems with Applications*, 112, 138–147. 10.1016/j.eswa.2018.06.027

Zee, P. C., & Goldstein, C. A. (2010). Treatment of shift work disorder and jet lag. *Current Treatment Options in Neurology*, 12, 396–411. doi: 10.1007/s11940-010-0090-9

Zhang, L., Sun, D. M., Li, C. B., & Tao, M. F. (2016). Influencing factors for sleep quality among shift-working nurses: A cross-sectional study in China using 3-factor pittsburgh sleep quality index. *Asian Nursing Research*, 10, 277–282. doi: 10.1016/j.anr.2016.09.002

Zhou, X., Ferguson, S. A., Matthews, R. W., Sargent, C., Darwent, D., Kennaway, D. J., & Roach, G. D. (2012). Mismatch between subjective alertness and objective performance under sleep restriction is greatest during the biological night. *Journal of Sleep Research*, 21, 40–49. doi: 10.1111/j.1365-2869.2011.00924.x

Zoer, I., Ruitenburg, M. M., Botje, D., Frings–Dresen, M. H. W., & Sluiter, J. K. (2011). The associations between psychosocial workload and mental health complaints in different age groups. *Ergonomics*, 54, 943–952. doi: 10.1080/00140139.2011.606920

Appendix A

INFORMED CONSENT

MORNING / AFTERNOON

I have been invited to participate in a research study using mobile devices that involves completing two-parts. The MORNING session needs to be completed in the morning (before 11:00) and the AFTERNOON session needs to be completed in the afternoon (after 15:00). Both the MORNING session and AFTERNOON session needs to be completed on consecutive days and will each take between 25 – 30 minutes to complete. Therefore, both the MORNING session and AFTERNOON session combined will take approximately 45 – 60 minutes.

I understand that the participation in this study is entirely voluntary and I may withdraw from the study at any time without giving any reason.

I understand that I am free to avoid responding to any question that I feel uncomfortable answering and that I can discuss my concerns with Professor Andy Smith via the email address mentioned below.

I understand that the information I am providing will be held totally anonymous, confidential and intended for academic research only, so that it is impossible to trace this information back to me individually. I understand that this information may be retained indefinitely.

I also understand that at the end of the study I will be provided with additional information and feedback about the purpose of the study.

By checking the box below and continuing, I consent to participate in the study conducted by Mr Michael Scott Evans (Doctoral Candidate), School of Psychology, Cardiff University, Wales, United Kingdom under the supervision of Professor Andy Smith.

Professor Andy Smith
School of Psychology
Cardiff University
63 Park Place
Cardiff
CF10 3AS
Tel: 029 2087 4757
Email: smithap@cf.ac.uk

Full name:____

Signed: ____

Date: ____

Appendix B

INSTRUCTIONS

MORNING SESSION

Thank you for agreeing to participate in this study. The aim of the study is to examine whether fatigue is associated with time of day. This will be examined using the Psychomotor Vigilance Task, which will be administered using mobile devices, such as those manufactured by both Apple and Samsung.

Please ensure you **DO NOT DRINK** any **CAFFEINE** (e.g., coffee, energy drinks, tea, etc.) and **ALCOHOL** beverages 24 hours prior to taking part in this study.

You will be asked to complete a two-part study on consecutive days using a mobile device. These two-parts will be the MORNING session (2 credits), which needs to be completed in the morning before 11:00 and the AFTERNOON session (2 credits), which needs to be completed in the afternoon after 15:00. A total of 4 credits will be awarded for this study through EMS, upon successful completion of the AFTERNOON session. Participation of this study will take approximately 25 – 30 minutes. Both the MORNING session and AFTERNOON session combined will take approximately 45 – 60 minutes to complete.

Participation of the MORNING session (2 credits) will involve completing:

1. A reaction time task (i.e., the Psychomotor Vigilance Task) (duration: approximately 10 minutes).

We would like to request that you be as open and honest as possible with your responses and avoid any perceptions of what you think a desired answer might be. The reliability of the data depends on your complete honesty. Please simply give the answer according to your opinion and your situation. Please try to make sure you have not inadvertently missed out any questions. For the Psychomotor Vigilance Task, please respond as quickly as possible.

Finally, we remind you that you are free to withdraw from the study at any point and if you feel uncomfortable answering any of the questions, you are free to not respond to those questions.

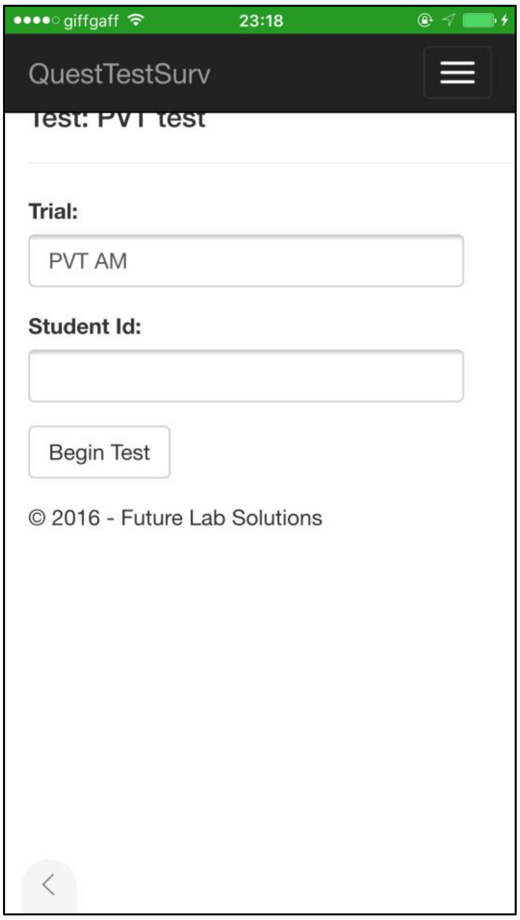
Thank you again for your participation.

MORNING User Guide for the Dolphin App

Step 1:
Load the Dolphin App **FIRST**.

Step 2:
Before you click on the **'Begin Test'** button, you will need to enter the following information in the boxes provided:

Trial: PVT AM
Student id: [Your Student ID]



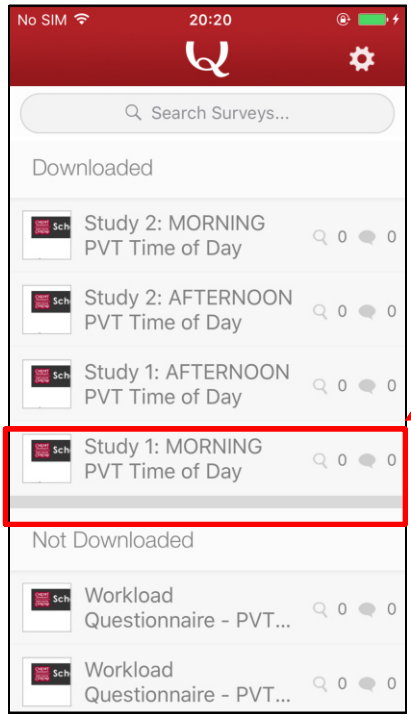
Step 3:
Follow on screen instructions.

Appendix D

Morning user guide for the Qualtrics App

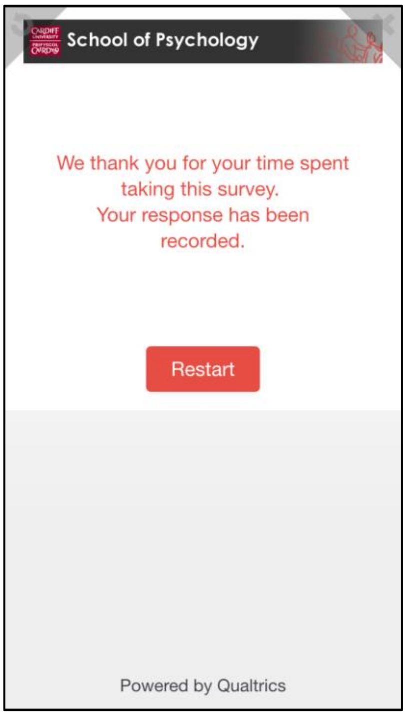
Step 1:
Once you have completed the Reaction Time Study using the Dolphin app, Load the Qualtrics App.

Step 2:
Click on **‘Study 1: MORNING PVT Time of Day’** survey



Load:
Study 1: MORNING PVT Time of Day

Step 3:
You have finished the survey when you see the following screen.



Appendix E

INSTRUCTIONS

AFTERNOON SESSION

Thank you for agreeing to participate in this study. The aim of the study is to examine whether fatigue is associated with time of day. This will be examined using the Psychomotor Vigilance Task, which will be administered using mobile devices, such as those manufactured by both Apple and Samsung.

Please ensure you **DO NOT DRINK** any **CAFFEINE** (e.g., coffee, energy drinks, tea, etc.) and **ALCOHOL** beverages 24 hours prior to taking part in this study.

You will be asked to complete a two-part study on consecutive days using a mobile device. These two-parts will be the MORNING session (2 credits), which needs to be completed in the morning before 11:00 and the AFTERNOON session (2 credits), which needs to be completed in the afternoon after 15:00. A total of 4 credits will be awarded for this study through EMS, upon successful completion of the AFTERNOON session. Participation of this study will take approximately 25 – 30 minutes. Both the MORNING session and AFTERNOON session combined will take approximately 45 – 60 minutes to complete.

Participation of the AFTERNOON session (2 credits) will involve completing:

1. A reaction time task (i.e., the Psychomotor Vigilance Task) (duration: approximately 10 minutes).
2. A 12-item questionnaire on your demographic information (duration: 1 – 3 minutes).

We would like to request that you be as open and honest as possible with your responses and avoid any perceptions of what you think a desired answer might be. The reliability of the data depends on your complete honesty. Please simply give the answer according to your opinion and your situation. Please try to make sure you have not inadvertently missed out any questions. For the Psychomotor Vigilance Task, please respond as quickly as possible.

Finally, we remind you that you are free to withdraw from the study at any point and if you feel uncomfortable answering any of the questions, you are free to not respond to those questions.

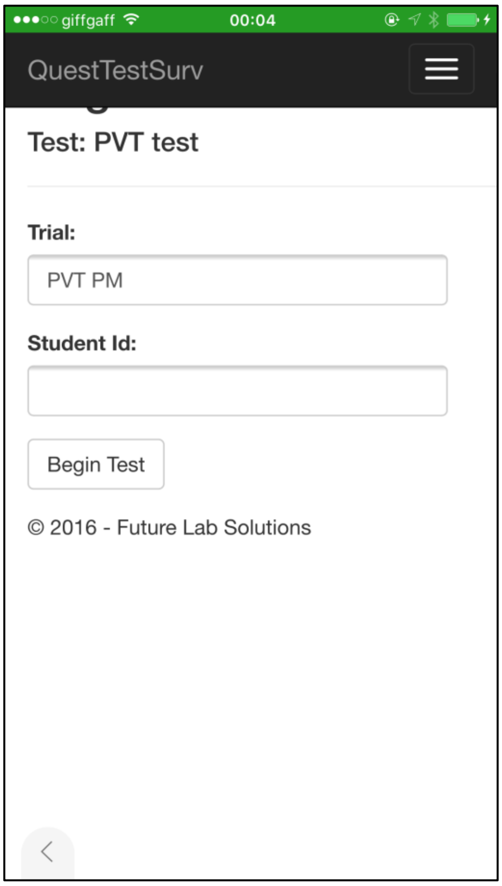
Thank you again for your participation.

AFTERNOON User Guide for the Dolphin App

Step 1:
Load the Dolphin App FIRST.

Step 2:
Before you click on the **'Begin Test'** button, you will need to enter the following information in the boxes provided:

Trial: PVT PM
Student id: [Your Student ID]

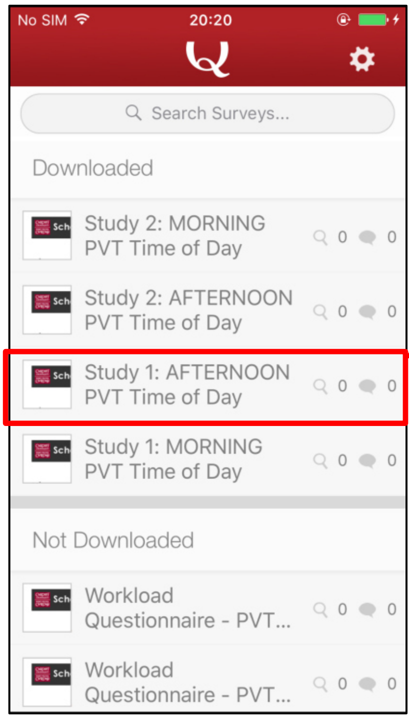


Step 3:
Follow on screen instructions.

Afternoon user guide for the Qualtrics App

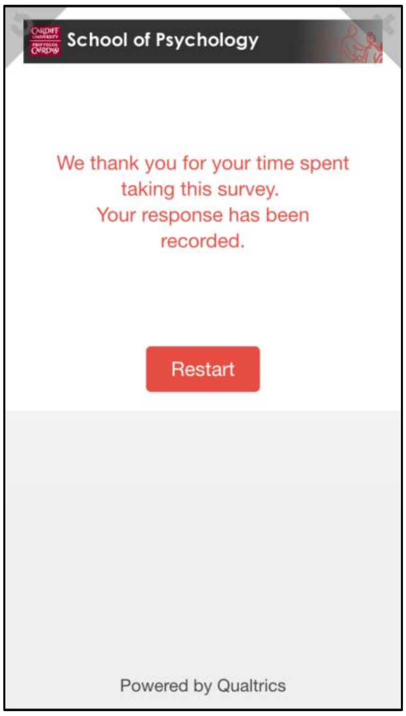
Step 1:
Once you have completed the Reaction Time Study using the Dolphin app, Load the Qualtrics App.

Step 2:
Click on **'Study 1: AFTERNOON PVT Time of Day'** survey



Load:
Study 1: AFTERNOON PVT Time of Day

Step 3:
You have finished the survey when you see the following screen.



Appendix H

DEBRIEF

Thank you for participating and completing both the MORNING session and AFTERNOON session of the Time of Day study, implemented on a mobile device. Using the Psychomotor Vigilance Task, the aim of this study is to examine whether fatigue is associated with time of day. In other words, whether cognitive levels of performance are associated with time of day. 4 Credits will be provided to you for participating in this study.

Each participant's electronic consent to participate and corresponding data that you have provided will be held completely anonymous, and only intended for academic research, so that it is impossible to trace this information back to any individual. Only Professor Andy Smith and the researcher (Mr Michael Scott Evans) will have access to your data. This information will be stored and analysed for publication before it is destroyed, in accordance with the Data Protection Act 1998.

If you have any queries or concerns about the research, please contact either the researcher Michael Scott Evans, or the supervisor Professor Andy Smith by using the contact details below. If you are affected by any of the issues raised in the four online tasks, or from the sleep quality questionnaire, then there are a number of services available through the university that is able to offer support, which can be accessed using the following links.

Staff of Centre for Occupational and Health Psychology
Equality and diversity
http://www.cardiff.ac.uk/govrn/cocom/equalityanddiversity/index.html

Counselling service
http://www.cardiff.ac.uk/counselling/about/index.html

If you still remain unhappy with the support and wish to complain formally, you can do so via the Cardiff University Ethics Committee:

Secretary of the Ethics Committee
School of Psychology
Cardiff University
Tower Building
Park Place
Cardiff
CF10 3AT
Tel: 029 2087 0360
Email: psychethics@cardiff.ac.uk

In addition, if you like to get any general information of the research findings such as published research articles and journals, please contact Michael Scott Evans (Doctoral candidate).

Thank you for your participation.

Mr. Michael Scott Evans
Doctoral Candidate
Centre for Occupational and Health Psychology
Cardiff School of Psychology
63 Park Place, Cardiff, Wales
United Kingdom, CF10 3AS
Email: EvansMS3@cardiff.ac.uk

Professor Andy Smith
Director/Supervisor
Centre for Occupational and Health Psychology
Cardiff School of Psychology
63 Park Place, Cardiff, Wales
United Kingdom, CF10 3AS
Email: smithap@cardiff.ac.uk

Appendix I

INFORMED CONSENT

AFTERNOON / MORNING

I have been invited to participate in a research study using mobile devices that involves completing two-parts. The AFTERNOON session needs to be completed in the afternoon (after 15:00) and the MORNING session needs to be completed in the morning (before 11:00). Both the AFTERNOON session and MORNING session needs to be completed on consecutive days and will each take between 25 – 30 minutes to complete. Therefore, both the AFTERNOON session and MORNING session combined will take approximately 45 – 60 minutes.

I understand that the participation in this study is entirely voluntary and I may withdraw from the study at any time without giving any reason.

I understand that I am free to avoid responding to any question that I feel uncomfortable answering and that I can discuss my concerns with Professor Andy Smith via the email address mentioned below.

I understand that the information I am providing will be held totally anonymous, confidential and intended for academic research only, so that it is impossible to trace this information back to me individually. I understand that this information may be retained indefinitely.

I also understand that at the end of the study I will be provided with additional information and feedback about the purpose of the study.

By checking the box below and continuing, I consent to participate in the study conducted by Mr Michael Scott Evans (Doctoral Candidate), School of Psychology, Cardiff University, Wales, United Kingdom under the supervision of Professor Andy Smith.

Professor Andy Smith
School of Psychology
Cardiff University
63 Park Place
Cardiff
CF10 3AS
Tel: 029 2087 4757
Email: smithap@cf.ac.uk

Full name:____

Signed: ____

Date: ____

Appendix J

INSTRUCTIONS

AFTERNOON SESSION

Thank you for agreeing to participate in this study. The aim of the study is to examine whether fatigue is associated with time of day. This will be examined using the Psychomotor Vigilance Task, which will be administered using mobile devices, such as those manufactured by both Apple and Samsung.

Please ensure you **DO NOT DRINK** any **CAFFEINE** (e.g., coffee, energy drinks, tea, etc.) and **ALCOHOL** beverages 24 hours prior to taking part in this study.

You will be asked to complete a two-part study on consecutive days using a mobile device. These two-parts will be the AFTERNOON session (2 credits), which needs to be completed in the morning before 11:00 and the MORNING session (2 credits), which needs to be completed in the afternoon after 15:00. A total of 4 credits will be awarded for this study through EMS, upon successful completion of the MORNING session. Participation of this study will take approximately 25 – 30 minutes. Both the AFTERNOON session and MORNING session combined will take approximately 45 – 60 minutes to complete.

Participation of the AFTERNOON session (2 credits) will involve completing:

1. A reaction time task (i.e., the Psychomotor Vigilance Task) (duration: approximately 10 minutes).

We would like to request that you be as open and honest as possible with your responses and avoid any perceptions of what you think a desired answer might be. The reliability of the data depends on your complete honesty. Please simply give the answer according to your opinion and your situation. Please try to make sure you have not inadvertently missed out any questions. For the Psychomotor Vigilance Task, please respond as quickly as possible.

Finally, we remind you that you are free to withdraw from the study at any point and if you feel uncomfortable answering any of the questions, you are free to not respond to those questions.

Thank you again for your participation.

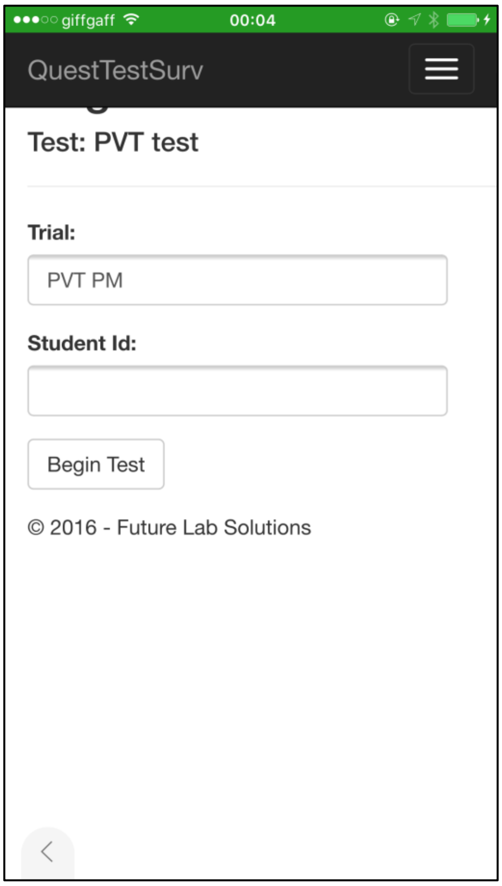
Appendix K

AFTERNOON User Guide for the Dolphin App

Step 1:
Load the Dolphin App FIRST.

Step 2:
Before you click on the **'Begin Test'** button, you will need to enter the following information in the boxes provided:

Trial: PVT PM
Student id: [Your Student ID]



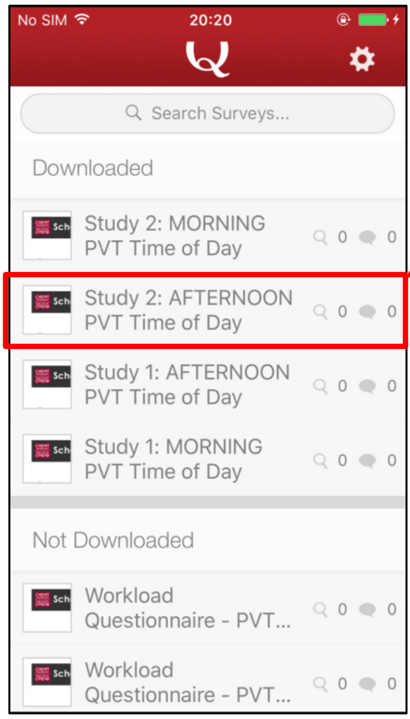
Step 3:
Follow on screen instructions.

Appendix L

Afternoon user guide for the Qualtrics App

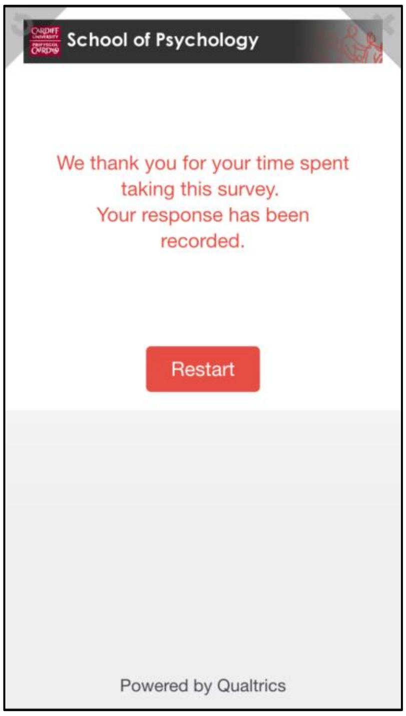
Step 1:
Once you have completed the Reaction Time Study using the Dolphin app, Load the Qualtrics App.

Step 2:
Click on **‘Study 2: AFTERNOON PVT Time of Day’** survey



Load:
Study 2: AFTERNOON PVT Time of Day

Step 3:
You have finished the survey when you see the following screen.



Appendix M

INSTRUCTIONS

MORNING SESSION

Thank you for agreeing to participate in this study. The aim of the study is to examine whether fatigue is associated with time of day. This will be examined using the Psychomotor Vigilance Task, which will be administered using mobile devices, such as those manufactured by both Apple and Samsung.

Please ensure you **DO NOT DRINK** any **CAFFEINE** (e.g., coffee, energy drinks, tea, etc.) and **ALCOHOL** beverages 24 hours prior to taking part in this study.

You will be asked to complete a two-part study on consecutive days using a mobile device. These two-parts will be the AFTERNOON session (2 credits), which needs to be completed in the morning before 11:00 and the MORNING session (2 credits), which needs to be completed in the afternoon after 15:00. A total of 4 credits will be awarded for this study through EMS, upon successful completion of the MORNING session. Participation of this study will take approximately 25 – 30 minutes. Both the AFTERNOON session and MORNING session combined will take approximately 45 – 60 minutes to complete.

Participation of the MORNING session (2 credits) will involve completing:

1. A reaction time task (i.e., the Psychomotor Vigilance Task) (duration: approximately 10 minutes).
2. A 12-item questionnaire on your demographic information (duration: 1 – 3 minutes).

We would like to request that you be as open and honest as possible with your responses and avoid any perceptions of what you think a desired answer might be. The reliability of the data depends on your complete honesty. Please simply give the answer according to your opinion and your situation. Please try to make sure you have not inadvertently missed out any questions. For the Psychomotor Vigilance Task, please respond as quickly as possible.

Finally, we remind you that you are free to withdraw from the study at any point and if you feel uncomfortable answering any of the questions, you are free to not respond to those questions.

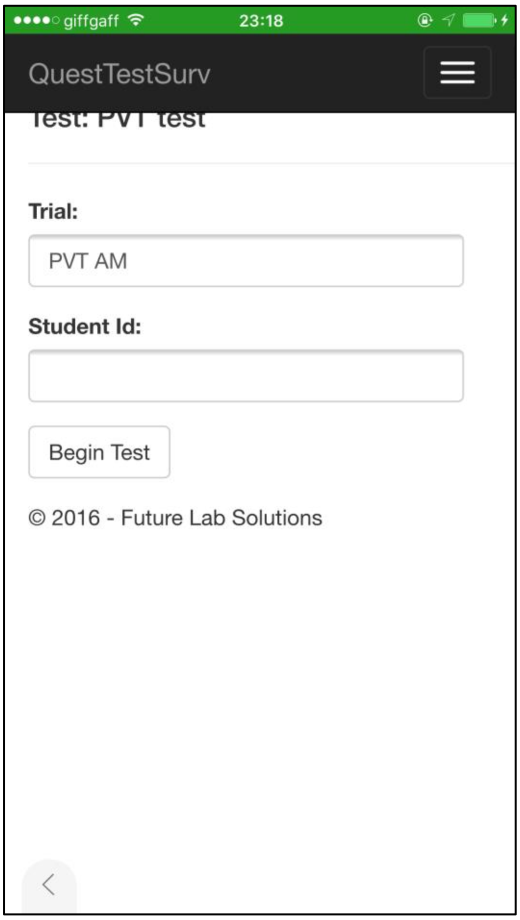
Thank you again for your participation.

MORNING User Guide for the Dolphin App

Step 1:
Load the Dolphin App **FIRST**.

Step 2:
Before you click on the **'Begin Test'** button, you will need to enter the following information in the boxes provided:

Trial: PVT AM
Student id: [Your Student ID]



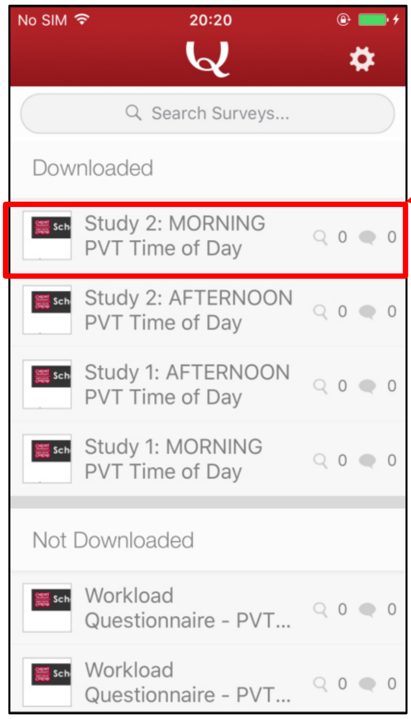
Step 3:
Follow on screen instructions.

Appendix O

Morning user guide for the Qualtrics App

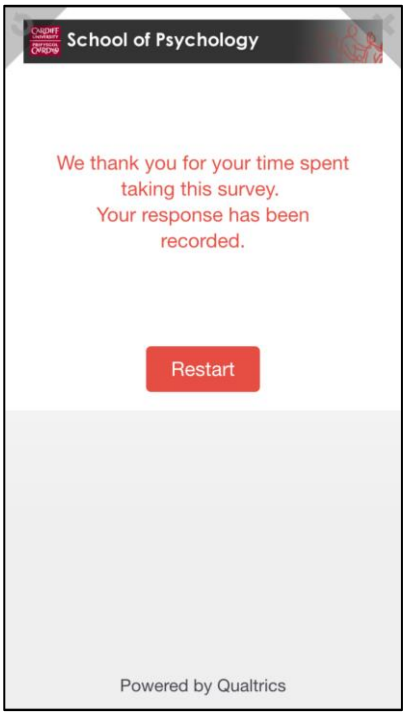
Step 1:
Once you have completed the Reaction Time Study using the Dolphin app, Load the Qualtrics App.

Step 2:
Click on **'Study 2: MORNING PVT Time of Day'** survey



Load:
Study 2: MORNING PVT Time of Day

Step 3:
You have finished the survey when you see the following screen.



Appendix P

DEBRIEF

Thank you for participating and completing both the AFTERNOON session and MORNING session of the Time of Day study, implemented on a mobile device. Using the Psychomotor Vigilance Task, the aim of this study is to examine whether fatigue is associated with time of day. In other words, whether cognitive levels of performance are associated with time of day. 4 Credits will be provided to you for participating in this study.

Each participant's electronic consent to participate and corresponding data that you have provided will be held completely anonymous, and only intended for academic research, so that it is impossible to trace this information back to any individual. Only Professor Andy Smith and the researcher (Mr Michael Scott Evans) will have access to your data. This information will be stored and analysed for publication before it is destroyed, in accordance with the Data Protection Act 1998.

If you have any queries or concerns about the research, please contact either the researcher Michael Scott Evans, or the supervisor Professor Andy Smith by using the contact details below. If you are affected by any of the issues raised in the four online tasks, or from the sleep quality questionnaire, then there are a number of services available through the university that is able to offer support, which can be accessed using the following links.

Staff of Centre for Occupational and Health Psychology
Equality and diversity
http://www.cardiff.ac.uk/govrn/cocom/equalityanddiversity/index.html

Counselling service
http://www.cardiff.ac.uk/counselling/about/index.html

If you still remain unhappy with the support and wish to complain formally, you can do so via the Cardiff University Ethics Committee:

Secretary of the Ethics Committee
School of Psychology
Cardiff University
Tower Building
Park Place
Cardiff
CF10 3AT
Tel: 029 2087 0360
Email: psychethics@cardiff.ac.uk

In addition, if you like to get any general information of the research findings such as published research articles and journals, please contact Michael Scott Evans (Doctoral candidate).

Thank you for your participation.

Mr. Michael Scott Evans
Doctoral Candidate
Centre for Occupational and Health Psychology
Cardiff School of Psychology
63 Park Place, Cardiff, Wales
United Kingdom, CF10 3AS
Email: EvansMS3@cardiff.ac.uk

Professor Andy Smith
Director/Supervisor
Centre for Occupational and Health Psychology
Cardiff School of Psychology
63 Park Place, Cardiff, Wales
United Kingdom, CF10 3AS
Email: smithap@cardiff.ac.uk

Appendix Q

INFORMED CONSENT

HIGH SIMULATED WORKLOAD

I have been invited to participate in a research study that involves completing two-parts on two different days; HIGH simulated workload and LOW simulated workload using a mobile device. Both the HIGH simulated workload session and LOW simulated workload session will each take between 55 – 60 minutes to complete. Therefore, when combining the HIGH simulated workload session and LOW simulated workload session, this study will take approximately 110 – 120 minutes to complete.

I understand that the participation in this study is entirely voluntary and I may withdraw from the study at any time without giving any reason.

I understand that I am free to avoid responding to any question that I feel uncomfortable answering and that I can discuss my concerns with Professor Andy Smith via the email address mentioned below.

I understand that the information I am providing will be held totally anonymous, confidential and intended for academic research only, so that it is impossible to trace this information back to me individually. I understand that this information may be retained indefinitely.

I also understand that at the end of the study I will be provided with additional information and feedback about the purpose of the study.

By checking the box below and continuing, I consent to participate in the study conducted by Mr Michael Scott Evans (Doctoral Candidate), School of Psychology, Cardiff University, Wales, United Kingdom under the supervision of Professor Andy Smith.

Professor Andy Smith
School of Psychology
Cardiff University
63 Park Place
Cardiff
CF10 3AS
Tel: 029 2087 4757
Email: smithap@cf.ac.uk

Full name:

____

Signed:

____

Date:

____

Appendix R

INSTRUCTIONS

HIGH SIMULATED WORKLOAD

Thank you for agreeing to participate in this study. The aim of the study is to examine whether fatigue is associated with simulated workload. This will be examined through a series of cognitive tasks, which will be administered remotely using Qualtrics an online external software survey platform, as well as using the using the Psychomotor Vigilance Task, which will be administered using mobile devices, such as those manufactured by both Apple and Samsung.

Please ensure you **DO NOT DRINK** any **CAFFEINE** (e.g., coffee, energy drinks, tea, etc.) and **ALCOHOL** beverages 24 hours prior to taking part in this study.

You are required to complete a two-part study using a mobile device. These are divided into two different days, a HIGH simulated workload day (5 credits) and a LOW simulated workload day (5 credits). A total of 10 credits will be awarded for this study through EMS or £20 upon successful completion of ALL two-parts. Participation of this study will take approximately 55 – 60 minutes. All two-parts combined will take approximately 110 – 120 minutes to complete.

This is the **HIGH** simulated workload day session. Participation of this session will involve completing:

1. A reaction time task (i.e., the Psychomotor Vigilance Task) (duration: 5 minutes).
2. Cognitive performance tasks:
   a. Logic processing task (duration: 8 minutes).
   b. Semantic word processing task (duration: 8 minutes).
   c. Semantic picture processing task (duration: 8 minutes).
   d. Serial search test (duration: 8 minutes).
3. A reaction time task (i.e., the Psychomotor Vigilance Task) (duration: 5 minutes).

We would like to request that you be as open and honest as possible with your responses and avoid any perceptions of what you think a desired answer might be. The reliability of the data depends on your complete honesty. Please simply give the answer according to your opinion and your situation. Please try to make sure you have not inadvertently missed out any questions. For the cognitive performance task as well as the psychomotor vigilance task, please respond as quickly and as accurately as possible.

Finally, we remind you that you are free to withdraw from the study at any point and if you feel uncomfortable answering any of the questions, you are free to not respond to those questions.

Thank you again for your participation.

Appendix S

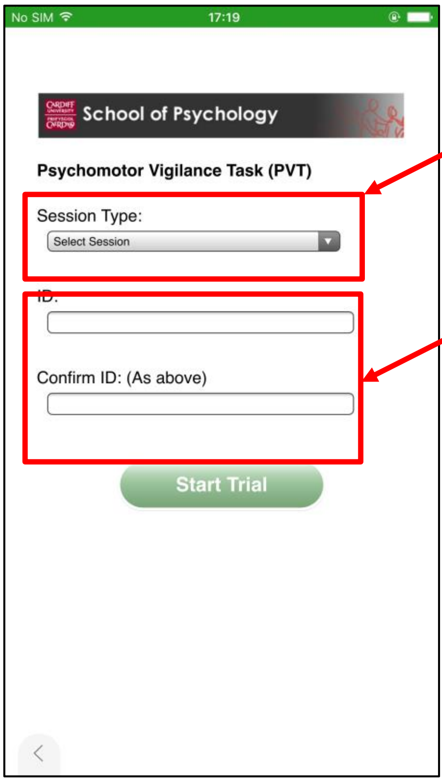
Reaction Time Study
User Guide for the Dolphin App

1

Step 1:
Load the Dolphin app.

Step 2:
Before you click on the **'Start Trial'** button, you will need to enter the following information in the boxes provided:

Session Type: As instructed by the experimenter
ID: [Your Student ID, e.g., 1573709]
Confirm ID (As above) (Re-enter your Student ID, e.g., 1573709]



Session Type:
Please select as instructed.

ID:
Please enter your student ID (e.g., 1573709).
Careful not to make a mistake.

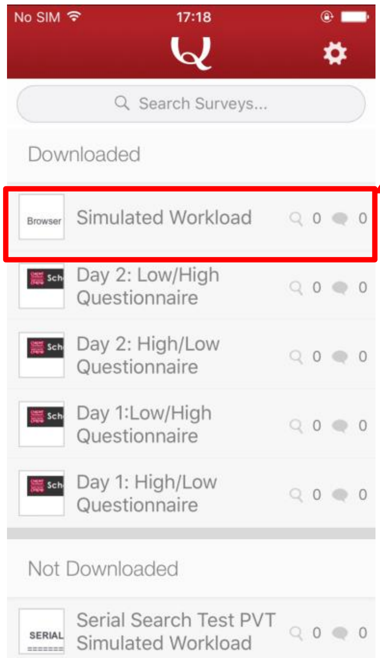
Step 3:
Follow on-screen instructions.

2

Cognitive Performance Tasks (Simulated Workload)

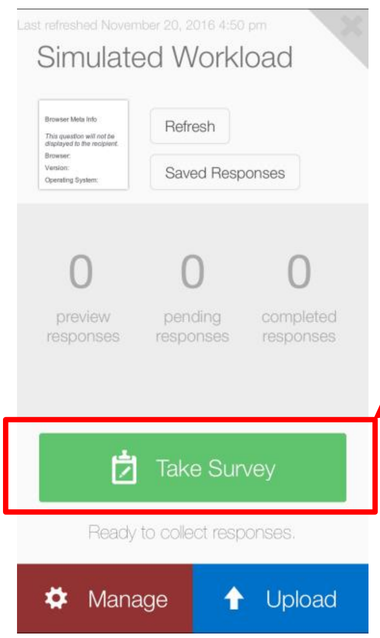
Step 1:
Once you have completed the Reaction Time Study using the Dolphin app, Load the Qualtrics App.

Step 2:
Click on **'Simulated Workload'** survey



Load:
Simulated Workload

Step 3:
Then click on **'Take Survey'**



Take Survey:
Simulated Workload

Step 4:
Follow the on-screen instructions

Step 5:
You have finished the simulated workload section when you see the following on-screen message below. Please grab the experimenter's attention upon view of this screen

Step 6:
Please grab the experimenter's attention upon view of the above on-screen message

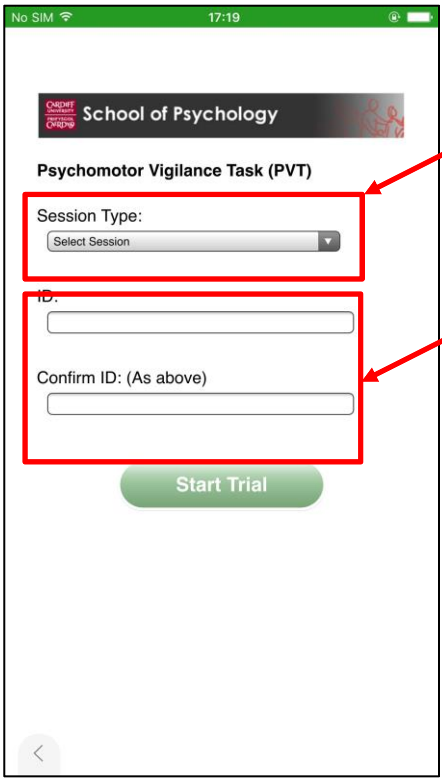
Reaction Time Study
User Guide for the Dolphin App

3

Step 1:
Load the Dolphin app.

Step 2:
Before you click on the **'Start Trial'** button, you will need to enter the following information in the boxes provided:

Session Type:	As instructed by the experimenter
ID:	[Your Student ID, e.g., 1573709]
Confirm ID (As above)	(Re-enter your Student ID, e.g., 1573709]



Session Type:
Please select as instructed.

ID:
Please enter your student ID (e.g., 1573709).
Careful not to make a mistake.

Step 3:
Follow on-screen instructions.

Appendix V

INFORMED CONSENT

LOW SIMULATED WORKLOAD

I have been invited to participate in a research study that involves completing two-parts on two different days; HIGH simulated workload and LOW simulated workload using a mobile device. Both the HIGH simulated workload session and LOW simulated workload session will each take between 55 – 60 minutes to complete. Therefore, when combining the HIGH simulated workload session and LOW simulated workload session, this study will take approximately 110 – 120 minutes to complete.

I understand that the participation in this study is entirely voluntary and I may withdraw from the study at any time without giving any reason.

I understand that I am free to avoid responding to any question that I feel uncomfortable answering and that I can discuss my concerns with Professor Andy Smith via the email address mentioned below.

I understand that the information I am providing will be held totally anonymous, confidential and intended for academic research only, so that it is impossible to trace this information back to me individually. I understand that this information may be retained indefinitely.

I also understand that at the end of the study I will be provided with additional information and feedback about the purpose of the study.

By checking the box below and continuing, I consent to participate in the study conducted by Mr Michael Scott Evans (Doctoral Candidate), School of Psychology, Cardiff University, Wales, United Kingdom under the supervision of Professor Andy Smith.

Professor Andy Smith
School of Psychology
Cardiff University
63 Park Place
Cardiff
CF10 3AS
Tel: 029 2087 4757
Email: smithap@cf.ac.uk

Full name:

____

Signed:

____

Date:

____

Appendix W

INSTRUCTIONS

LOW SIMULATED WORKLOAD

Thank you for agreeing to participate in this study. The aim of the study is to examine whether fatigue is associated with simulated workload. This will be examined through a series of cognitive tasks, which will be administered remotely using Qualtrics an online external software survey platform, as well as using the using the Psychomotor Vigilance Task, which will be administered using mobile devices, such as those manufactured by both Apple and Samsung.

Please ensure you **DO NOT DRINK** any **CAFFEINE** (e.g., coffee, energy drinks, tea, etc.) and **ALCOHOL** beverages 24 hours prior to taking part in this study.

You are required to complete a two-part study using a mobile device. These are divided into two different days, a HIGH simulated workload day (5 credits) and a LOW simulated workload day (5 credits). A total of 10 credits will be awarded for this study through EMS or £20 upon successful completion of ALL two-parts. Participation of this study will take approximately 55 – 60 minutes. All two-parts combined will take approximately 110 – 120 minutes to complete.

This is the **LOW** simulated workload day session. Participation of this session will involve completing:

1. A reaction time task (i.e., the Psychomotor Vigilance Task) (duration: 5 minutes).
2. Watching a television show e.g., The Big Bang Theory (duration 32 minutes).
3. A reaction time task (i.e., the Psychomotor Vigilance Task) (duration: 5 minutes).
4. Questionnaires
    a. A 12-item questionnaire on your demographic information (duration: 1 – 3 minutes).

We would like to request that you be as open and honest as possible with your responses and avoid any perceptions of what you think a desired answer might be. The reliability of the data depends on your complete honesty. Please simply give the answer according to your opinion and your situation. Please try to make sure you have not inadvertently missed out any questions. For the cognitive performance task as well as the psychomotor vigilance task, please respond as quickly and as accurately as possible.

Finally, we remind you that you are free to withdraw from the study at any point and if you feel uncomfortable answering any of the questions, you are free to not respond to those questions.

Thank you again for your participation.

Appendix X

Reaction Time Study

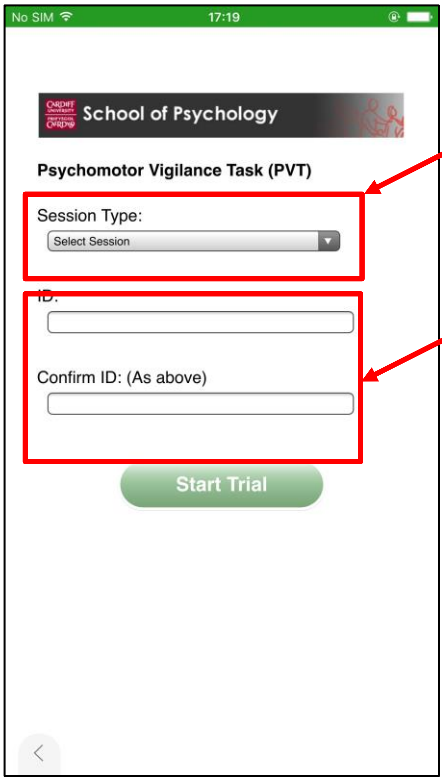
User Guide for the Dolphin App

1

Step 1:
Load the Dolphin app.

Step 2:
Before you click on the **‘Start Trial’** button, you will need to enter the following information in the boxes provided:

Session Type:	As instructed by the experimenter
ID:	[Your Student ID, e.g., 1573709]
Confirm ID (As above)	(Re-enter your Student ID, e.g., 1573709]



Session Type:
Please select as instructed.

ID:
Please enter your student ID (e.g., 1573709).
Careful not to make a mistake.

Step 3:
Follow on-screen instructions.

Television Show

2

Appendix Z

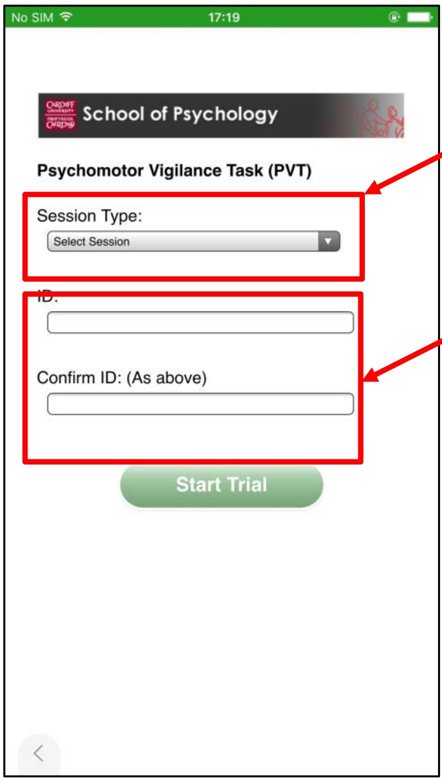
Reaction Time Study
User Guide for the Dolphin App

3

Step 1:
Load the Dolphin app.

Step 2:
Before you click on the **'Start Trial'** button, you will need to enter the following information in the boxes provided:

Session Type:	As instructed by the experimenter
ID:	[Your Student ID, e.g., 1573709]
Confirm ID (As above)	(Re-enter your Student ID, e.g., 1573709]



Session Type:
Please select as instructed.

ID:
Please enter your student ID (e.g., 1573709).
Careful not to make a mistake.

Step 3:
Follow on-screen instructions.

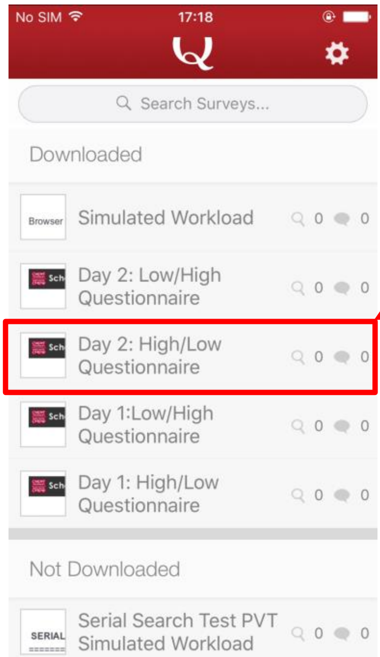
Appendix AA

Qualtrics: 'Day 2 High/Low Questionnaire'

4

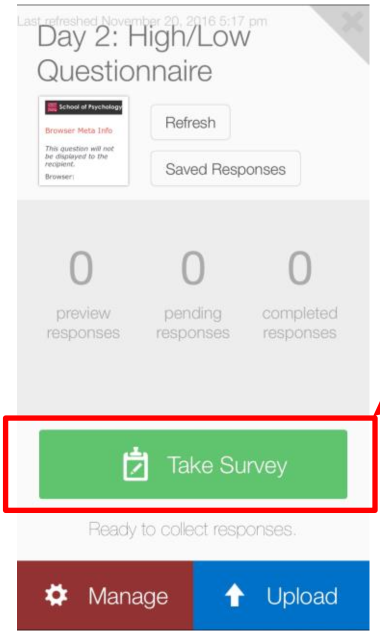
Step 1:
Once you have completed the second Reaction Time Study using the Dolphin app, Load the Qualtrics App once again.

Step 2:
Click on **'Day 2 High/Low Questionnaire'** survey



Load:
Day 2 High/Low Questionnaire

Step 3:
Then click on **'Take Survey'**



Take Survey:
Day 2 High/Low Questionnaire

Step 4:
Once you have completed the Questionnaire, please hand over the iPhone to the experimenter.

Appendix BB

DEBRIEF

Thank you for participating and completing all two-parts of the workload study, implemented on a mobile device. Using the cognitive performance tasks as well as the psychomotor vigilance task, the aim of this study is to examine whether fatigue is associated with simulated workload. In other words, whether cognitive levels of performance are associated with simulated workload. 10 Credits or £20 will be provided to you for participating in this study.

Each participant's electronic consent to participate and corresponding data that you have provided will be held completely anonymous, and only intended for academic research, so that it is impossible to trace this information back to any individual. Only Professor Andy Smith and the researcher (Mr Michael Scott Evans) will have access to your data. This information will be stored and analysed for publication before it is destroyed, in accordance with the Data Protection Act 1998.

If you have any queries or concerns about the research, please contact either the researcher Michael Scott Evans, or the supervisor Professor Andy Smith by using the contact details below. If you are affected by any of the issues raised in the four online tasks, or from the sleep quality questionnaire, then there are a number of services available through the university that is able to offer support, which can be accessed using the following links.

Staff of Centre for Occupational and Health Psychology
Equality and diversity
http://www.cardiff.ac.uk/govrn/cocom/equalityanddiversity/index.html

Counselling service
http://www.cardiff.ac.uk/counselling/about/index.html

If you still remain unhappy with the support and wish to complain formally, you can do so via the Cardiff University Ethics Committee:

Secretary of the Ethics Committee
School of Psychology
Cardiff University
Tower Building
Park Place
Cardiff
CF10 3AT
Tel: 029 2087 0360
Email: psychethics@cardiff.ac.uk

In addition, if you like to get any general information of the research findings such as published research articles and journals, please contact Michael Scott Evans (Doctoral candidate).

Thank you for your participation.

Mr. Michael Scott Evans
Doctoral Candidate
Centre for Occupational and Health Psychology
Cardiff School of Psychology
63 Park Place, Cardiff, Wales
United Kingdom, CF10 3AS
Email: EvansMS3@cardiff.ac.uk

Professor Andy Smith
Director/Supervisor
Centre for Occupational and Health Psychology
Cardiff School of Psychology
63 Park Place, Cardiff, Wales
United Kingdom, CF10 3AS
Email: smithap@cardiff.ac.uk

Appendix CC

INFORMED CONSENT

LOW SIMULATED WORKLOAD

I have been invited to participate in a research study that involves completing two-parts on two different days; LOW simulated workload and HIGH simulated workload using a mobile device. Both the LOW simulated workload session and HIGH simulated workload session will each take between 55 – 60 minutes to complete. Therefore, when combining the LOW simulated workload session and HIGH simulated workload session, this study will take approximately 110 – 120 minutes to complete.

I understand that the participation in this study is entirely voluntary and I may withdraw from the study at any time without giving any reason.

I understand that I am free to avoid responding to any question that I feel uncomfortable answering and that I can discuss my concerns with Professor Andy Smith via the email address mentioned below.

I understand that the information I am providing will be held totally anonymous, confidential and intended for academic research only, so that it is impossible to trace this information back to me individually. I understand that this information may be retained indefinitely.

I also understand that at the end of the study I will be provided with additional information and feedback about the purpose of the study.

By checking the box below and continuing, I consent to participate in the study conducted by Mr Michael Scott Evans (Doctoral Candidate), School of Psychology, Cardiff University, Wales, United Kingdom under the supervision of Professor Andy Smith.

Professor Andy Smith
School of Psychology
Cardiff University
63 Park Place
Cardiff
CF10 3AS
Tel: 029 2087 4757
Email: smithap@cf.ac.uk

Full name:

____

Signed:

____

Date:

____

Appendix DD

INSTRUCTIONS

LOW SIMULATED WORKLOAD

Thank you for agreeing to participate in this study. The aim of the study is to examine whether fatigue is associated with simulated workload. This will be examined through a series of cognitive tasks, which will be administered remotely using Qualtrics an online external software survey platform, as well as using the using the Psychomotor Vigilance Task, which will be administered using mobile devices, such as those manufactured by both Apple and Samsung.

Please ensure you **DO NOT DRINK** any **CAFFEINE** (e.g., coffee, energy drinks, tea, etc.) and **ALCOHOL** beverages 24 hours prior to taking part in this study.

You are required to complete a two-part study using a mobile device. These are divided into two different days, a LOW simulated workload day (5 credits) and a HIGH simulated workload day (5 credits). A total of 10 credits will be awarded for this study through EMS or £20 upon successful completion of ALL two-parts. Participation of this study will take approximately 55 – 60 minutes. All two-parts combined will take approximately 110 – 120 minutes to complete.

This is the **LOW** simulated workload day session. Participation of this session will involve completing:

1. A reaction time task (i.e., the Psychomotor Vigilance Task) (duration: 5 minutes).
2. Watching a television show e.g., The Big Bang Theory (duration 32 minutes).
3. A reaction time task (i.e., the Psychomotor Vigilance Task) (duration: 5 minutes).

We would like to request that you be as open and honest as possible with your responses and avoid any perceptions of what you think a desired answer might be. The reliability of the data depends on your complete honesty. Please simply give the answer according to your opinion and your situation. Please try to make sure you have not inadvertently missed out any questions. For the cognitive performance task as well as the psychomotor vigilance task, please respond as quickly and as accurately as possible.

Finally, we remind you that you are free to withdraw from the study at any point and if you feel uncomfortable answering any of the questions, you are free to not respond to those questions.

Thank you again for your participation.

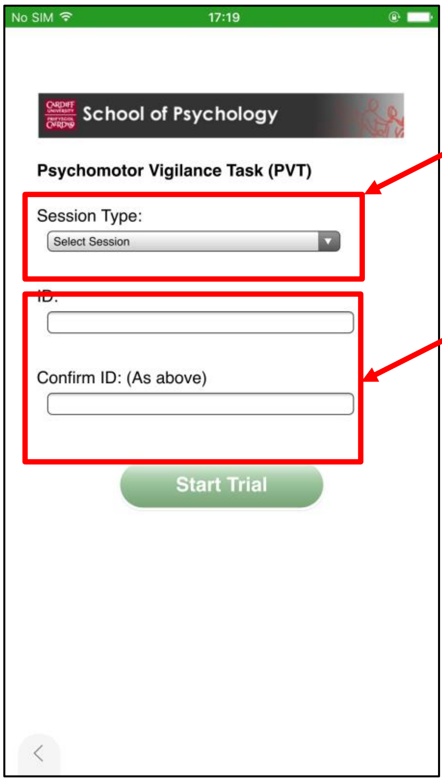
Reaction Time Study
User Guide for the Dolphin App

1

Step 1:
Load the Dolphin app.

Step 2:
Before you click on the **'Start Trial'** button, you will need to enter the following information in the boxes provided:

Session Type:	As instructed by the experimenter
ID:	[Your Student ID, e.g., 1573709]
Confirm ID (As above)	(Re-enter your Student ID, e.g., 1573709]



Session Type:
Please select as instructed.

ID:
Please enter your student ID (e.g., 1573709).
Careful not to make a mistake.

Step 3:
Follow on-screen instructions.

Television Show

2

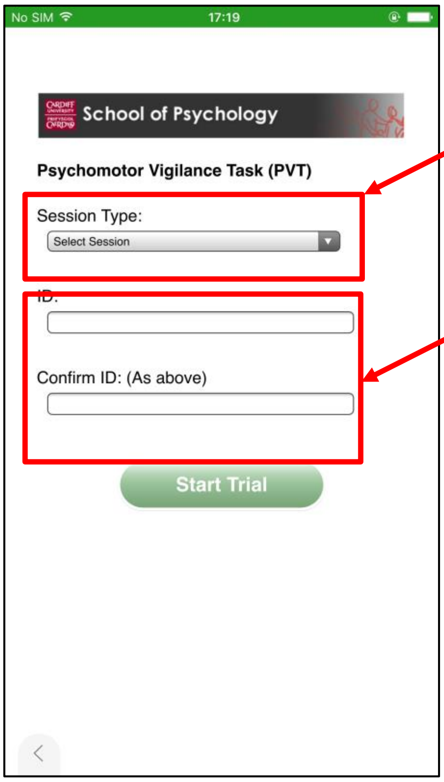
3

Reaction Time Study
User Guide for the Dolphin App

Step 1:
Load the Dolphin app.

Step 2:
Before you click on the **'Start Trial'** button, you will need to enter the following information in the boxes provided:

Session Type:	As instructed by the experimenter
ID:	[Your Student ID, e.g., 1573709]
Confirm ID (As above)	(Re-enter your Student ID, e.g., 1573709]



Session Type:
Please select as instructed.

ID:
Please enter your student ID (e.g., 1573709).
Careful not to make a mistake.

Step 3:
Follow on-screen instructions.

Appendix HH

INFORMED CONSENT

HIGH SIMULATED WORKLOAD

I have been invited to participate in a research study that involves completing two-parts on two different days; LOW simulated workload and HIGH simulated workload using a mobile device. Both the LOW simulated workload session and HIGH simulated workload session will each take between 55 – 60 minutes to complete. Therefore, when combining the LOW simulated workload session and HIGH simulated workload session, this study will take approximately 110 – 120 minutes to complete.

I understand that the participation in this study is entirely voluntary and I may withdraw from the study at any time without giving any reason.

I understand that I am free to avoid responding to any question that I feel uncomfortable answering and that I can discuss my concerns with Professor Andy Smith via the email address mentioned below.

I understand that the information I am providing will be held totally anonymous, confidential and intended for academic research only, so that it is impossible to trace this information back to me individually. I understand that this information may be retained indefinitely.

I also understand that at the end of the study I will be provided with additional information and feedback about the purpose of the study.

By checking the box below and continuing, I consent to participate in the study conducted by Mr Michael Scott Evans (Doctoral Candidate), School of Psychology, Cardiff University, Wales, United Kingdom under the supervision of Professor Andy Smith.

Professor Andy Smith
School of Psychology
Cardiff University
63 Park Place
Cardiff
CF10 3AS
Tel: 029 2087 4757
Email: smithap@cf.ac.uk

Full name:

____

Signed:

____

Date:

____

Appendix II

INSTRUCTIONS

HIGH SIMULATED WORKLOAD

Thank you for agreeing to participate in this study. The aim of the study is to examine whether fatigue is associated with simulated workload. This will be examined through a series of cognitive tasks, which will be administered remotely using Qualtrics an online external software survey platform, as well as using the using the Psychomotor Vigilance Task, which will be administered using mobile devices, such as those manufactured by both Apple and Samsung.

Please ensure you **DO NOT DRINK** any **CAFFEINE** (e.g., coffee, energy drinks, tea, etc.) and **ALCOHOL** beverages 24 hours prior to taking part in this study.

You are required to complete a two-part study using a mobile device. These are divided into two different days, a LOW simulated workload day (5 credits) and a HIGH simulated workload day (5 credits). A total of 10 credits will be awarded for this study through EMS or £20 upon successful completion of ALL two-parts. Participation of this study will take approximately 55 – 60 minutes. All two-parts combined will take approximately 110 – 120 minutes to complete.

This is the **HIGH** simulated workload day session. Participation of this session will involve completing:

1. A reaction time task (i.e., the Psychomotor Vigilance Task) (duration: 5 minutes).
2. Cognitive performance tasks:
    a. Logic processing task (duration: 8 minutes).
    b. Semantic word processing task (duration: 8 minutes).
    c. Semantic picture processing task (duration: 8 minutes).
    d. Serial search test (duration: 8 minutes).
3. A reaction time task (i.e., the Psychomotor Vigilance Task) (duration: 5 minutes).
4. Questionnaires
    a. A 12-item questionnaire on your demographic information (duration: 1 – 3 minutes).

We would like to request that you be as open and honest as possible with your responses and avoid any perceptions of what you think a desired answer might be. The reliability of the data depends on your complete honesty. Please simply give the answer according to your opinion and your situation. Please try to make sure you have not inadvertently missed out any questions. For the cognitive performance task as well as the psychomotor vigilance task, please respond as quickly and as accurately as possible.

Finally, we remind you that you are free to withdraw from the study at any point and if you feel uncomfortable answering any of the questions, you are free to not respond to those questions.

Thank you again for your participation.

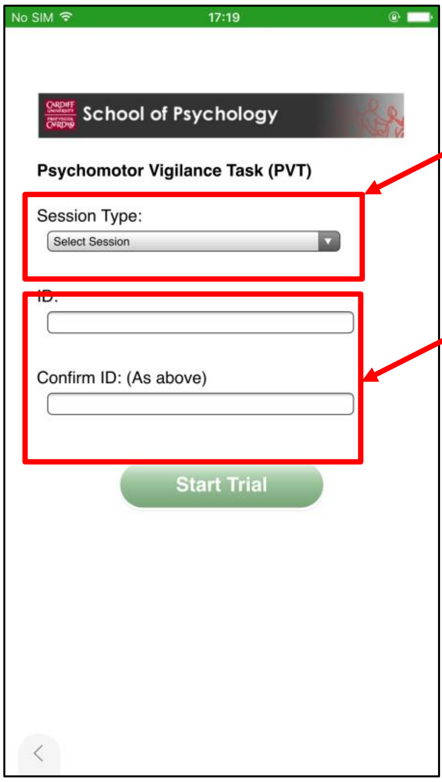
Reaction Time Study
User Guide for the Dolphin App

1

Step 1:
Load the Dolphin app.

Step 2:
Before you click on the **'Start Trial'** button, you will need to enter the following information in the boxes provided:

Session Type:	As instructed by the experimenter
ID:	[Your Student ID, e.g., 1573709]
Confirm ID (As above)	(Re-enter your Student ID, e.g., 1573709]



Session Type:
Please select as instructed.

ID:
Please enter your student ID (e.g., 1573709).
Careful not to make a mistake.

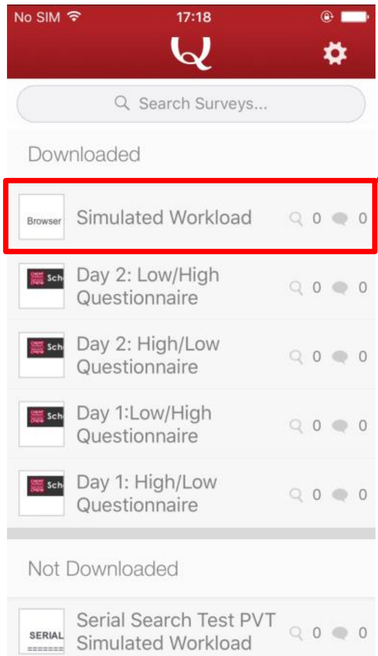
Step 3:
Follow on-screen instructions.

Cognitive Performance Tasks (Simulated Workload)

2

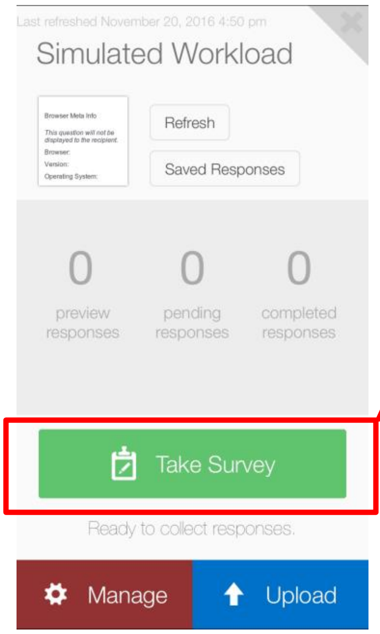
Step 1:
Once you have completed the Reaction Time Study using the Dolphin app, Load the Qualtrics App.

Step 2:
Click on **'Simulated Workload'** survey



Load:
Simulated Workload

Step 3:
Then click on **'Take Survey'**



Take Survey:
Simulated Workload

Step 4:
Follow the on-screen instructions

Step 5:
You have finished the simulated workload section when you see the following on-screen message below. Please grab the experimenter’s attention upon view of this screen

Step 6:
Please grab the experimenter’s attention upon view of the above on-screen message

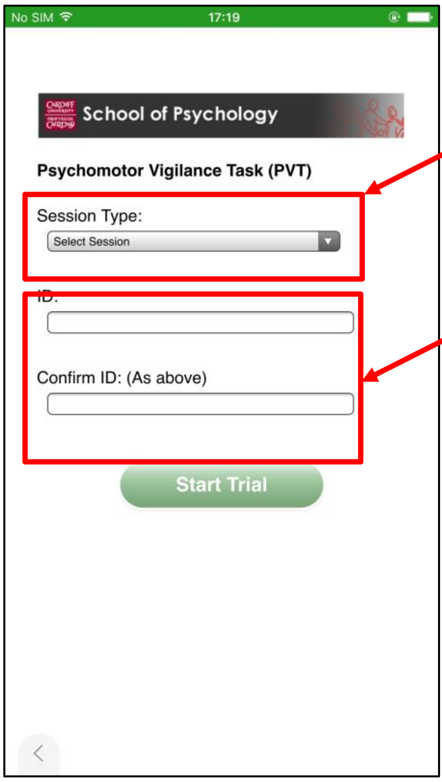
3

Reaction Time Study
User Guide for the Dolphin App

Step 1:
Load the Dolphin app.

Step 2:
Before you click on the **'Start Trial'** button, you will need to enter the following information in the boxes provided:

Session Type:	As instructed by the experimenter
ID:	[Your Student ID, e.g., 1573709]
Confirm ID (As above)	(Re-enter your Student ID, e.g., 1573709]



Session Type:
Please select as instructed.

ID:
Please enter your student ID (e.g., 1573709).
Careful not to make a mistake.

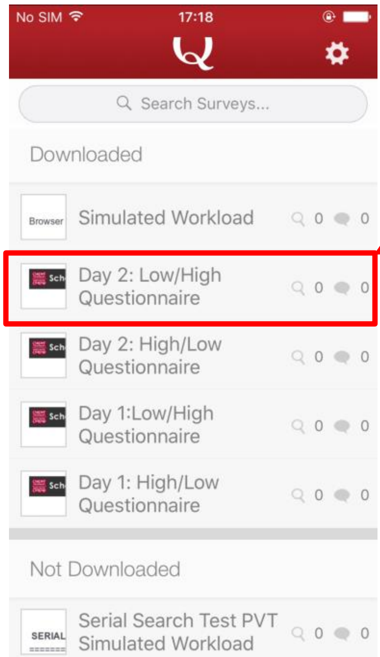
Step 3:
Follow on-screen instructions.

4

Qualtrics: ‘Day 2 Low/High Questionnaire’

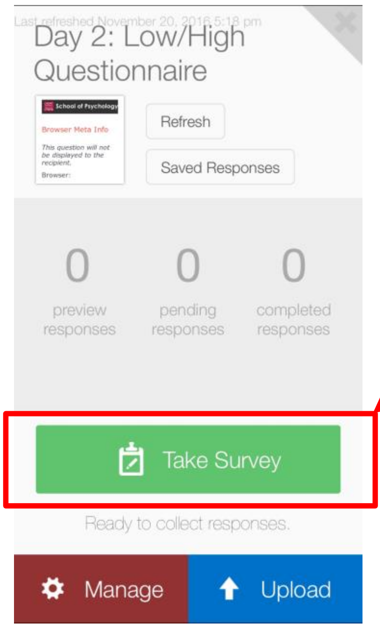
Step 1:
Once you have completed the second Reaction Time Study using the Dolphin app, Load the Qualtrics App once again.

Step 2:
Click on **‘Day 2 Low/High Questionnaire’** survey



Load:
Day 2 Low/High Questionnaire

Step 3:
Then click on **‘Take Survey’**



Take Survey:
Day 2 Low/High Questionnaire

Step 4:
Once you have completed the Questionnaire, please hand over the iPhone to the experimenter.

Appendix NN

DEBRIEF

Thank you for participating and completing all two-parts of the workload study, implemented on a mobile device. Using the cognitive performance tasks as well as the psychomotor vigilance task, the aim of this study is to examine whether fatigue is associated with simulated workload. In other words, whether cognitive levels of performance are associated with simulated workload. 10 Credits or £20 will be provided to you for participating in this study.

Each participant's electronic consent to participate and corresponding data that you have provided will be held completely anonymous, and only intended for academic research, so that it is impossible to trace this information back to any individual. Only Professor Andy Smith and the researcher (Mr Michael Scott Evans) will have access to your data. This information will be stored and analysed for publication before it is destroyed, in accordance with the Data Protection Act 1998.

If you have any queries or concerns about the research, please contact either the researcher Michael Scott Evans, or the supervisor Professor Andy Smith by using the contact details below. If you are affected by any of the issues raised in the four online tasks, or from the sleep quality questionnaire, then there are a number of services available through the university that is able to offer support, which can be accessed using the following links.

Staff of Centre for Occupational and Health Psychology
Equality and diversity
http://www.cardiff.ac.uk/govrn/cocom/equalityanddiversity/index.html

Counselling service
http://www.cardiff.ac.uk/counselling/about/index.html

If you still remain unhappy with the support and wish to complain formally, you can do so via the Cardiff University Ethics Committee:

Secretary of the Ethics Committee
School of Psychology
Cardiff University
Tower Building
Park Place
Cardiff
CF10 3AT
Tel: 029 2087 0360
Email: psychethics@cardiff.ac.uk

In addition, if you like to get any general information of the research findings such as published research articles and journals, please contact Michael Scott Evans (Doctoral candidate).

Thank you for your participation.

Mr. Michael Scott Evans
Doctoral Candidate
Centre for Occupational and Health Psychology
Cardiff School of Psychology
63 Park Place, Cardiff, Wales
United Kingdom, CF10 3AS
Email: EvansMS3@cardiff.ac.uk

Professor Andy Smith
Director/Supervisor
Centre for Occupational and Health Psychology
Cardiff School of Psychology
63 Park Place, Cardiff, Wales
United Kingdom, CF10 3AS
Email: smithap@cardiff.ac.uk

Appendix OO

INFORMED CONSENT

I have been invited to participate in a research study that involves completing four-parts on two different days; LOW workload (i.e., spare driver) and HIGH workload (i.e., a difficult diagram), at two different time points (BEFORE diagram and AFTER diagram) using a mobile device. A total of £20 will be awarded to participants, upon successful completion of all four-parts. All payments will be made via pre-written cheques. Participation of this study will take approximately 15 – 20 minutes per session. All four-sessions combined will take approximately 60 – 80 minutes to complete.

I understand that the participation in this study is entirely voluntary and I may withdraw from the study at any time without giving any reason.

I understand that I am free to avoid responding to any question that I feel uncomfortable answering and that I can discuss my concerns with Professor Andy Smith via the email address mentioned below.

I understand that the information I am providing will be held totally anonymous and intended for academic research only, so that it is impossible to trace this information back to me individually. I understand that this information may be retained indefinitely.

I also understand that at the end of the study I will be provided with additional information and feedback about the purpose of the study.

By signing below, I consent to participate in the study conducted by Mr Michael Scott Evans (Doctoral Candidate), School of Psychology, Cardiff University, Wales, United Kingdom under the supervision of Professor Andy Smith.

Professor Andy Smith
School of Psychology
Cardiff University
63 Park Place
Cardiff
CF10 3AS
Tel: 029 2087 4757
Email: smithap@cf.ac.uk

Full name:

___

Signed:

___

Date:

___

Appendix PP

INFORMATION SHEET

HIGH WORKLOAD BEFORE DIAGRAM

Thank you for agreeing to participate in this study. The aim of the study is to examine whether fatigue is associated with workload. This will be examined using the Psychomotor Vigilance Task, which will be administered using an Apple iPhone 6s Plus.

You will be kindly asked to complete a four-part study using a mobile device. These are divided into two different days, a LOW workload (i.e., spare driver) and HIGH workload (i.e., a difficult diagram), at two different time points (BEFORE diagram and AFTER diagram) using a mobile device. A total of £20 will be awarded to participants, upon successful completion of all four-parts. All payments will be made via pre-written cheques. Participation of this study will take approximately 15 – 20 minutes per session. All four-sessions combined will take approximately 60 – 80 minutes to complete.

This is the **HIGH** workload day **BEFORE** diagram session. Participation of this session will involve completing:

3. A reaction time task (duration: 10 minutes);
4. A 12-item questionnaire on your sleep quality (duration: approximately 1 - 3 minutes).

We would like to request that you be as open and honest as possible with your responses and avoid any perceptions of what you think a desired answer might be. The reliability of the data depends on your complete honesty for the sleep quality questionnaire. Please simply give the answer according to your opinion and your situation. Please try to make sure you have not inadvertently missed out any questions. For the reaction time task, please respond as quickly as possible.

Finally, we remind you that you are free to withdraw from the study at any point and if you feel uncomfortable answering any of the questions, you are free to not respond to those questions.

Thank you again for your participation.

ATW – Day 1: Session 1

1

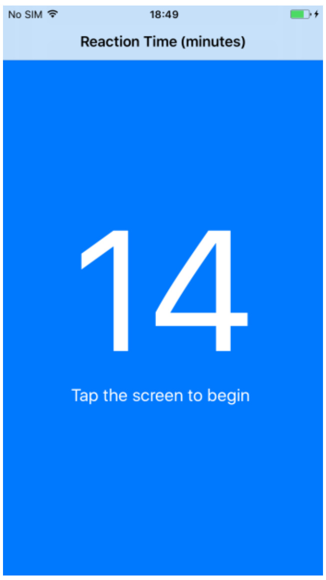
Step 1:
Launch the **Reaction Time** app.

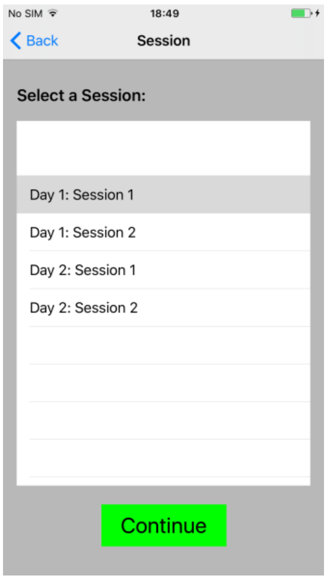
Step 2:
Enter both your **Volunteer Number** and **Password**, provided by the experimenter.

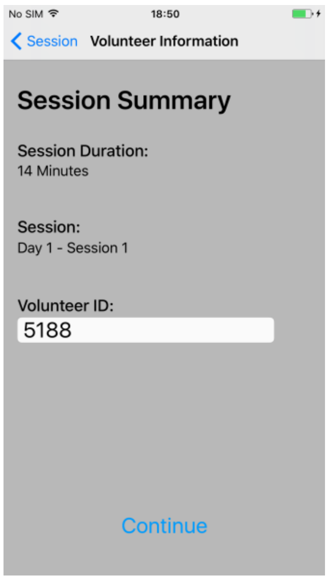
Enter:
Volunteer Number: 5188

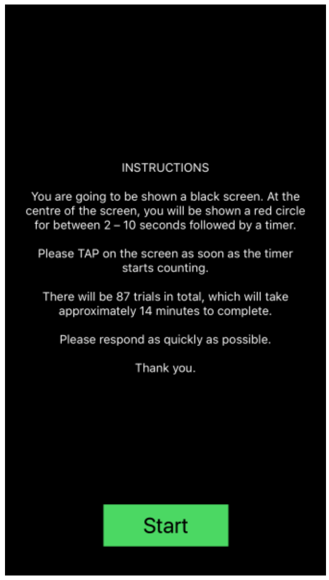
Enter:
Password: 504

Step 3: Tap anywhere on the screen to continue.	**Step 4:** Select **Day 1: Session 1**, then tap **Continue**.	**Step 5:** Make sure your **Session** and **Volunteer ID** are correct, then tap **Continue**.	**Step 6:** Follow the on-screen Instructions carefully and tap **Start** to begin.







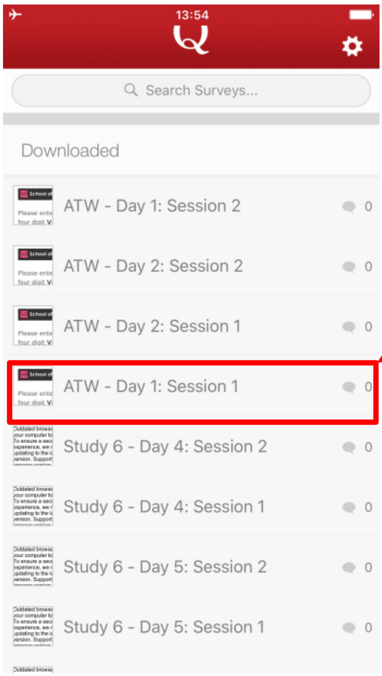


ATW – Day 1: Session 1

2

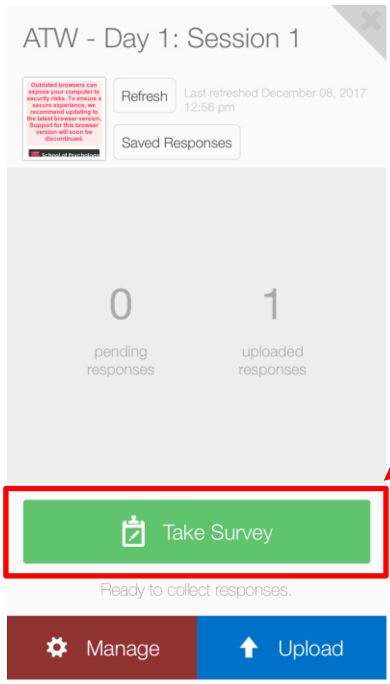
Step 1:
Launch the **Qualtrics Surveys** app.

Step 2:
Click on the **'ATW – Day 1: Session 1'** survey



Select:
ATW – Day 1: Session 1

Step 3:
Then tap on **'Take Survey'**



Tap:
Take Survey

Step 4:
Once you have completed the questionnaire, you are done!

Appendix SS

INFORMATION SHEET

HIGH WORKLOAD AFTER DIAGRAM

Thank you for agreeing to participate in this study. The aim of the study is to examine whether fatigue is associated with workload. This will be examined using the Psychomotor Vigilance Task, which will be administered using an Apple iPhone 6s Plus.

You will be kindly asked to complete a four-part study using a mobile device. These are divided into two different days, a LOW workload (i.e., spare driver) and HIGH workload (i.e., a difficult diagram), at two different time points (BEFORE diagram and AFTER diagram) using a mobile device. A total of £20 will be awarded to participants, upon successful completion of all four-parts. All payments will be made via pre-written cheques. Participation of this study will take approximately 15 – 20 minutes per session. All four-sessions combined will take approximately 60 – 80 minutes to complete.

This is the **HIGH** workload day **AFTER** diagram session. Participation of this session will involve completing:

1. A reaction time task (duration: 10 minutes);
2. An 8-item questionnaire on your workload (duration: approximately 1 minute).
3. A 10-item questionnaire on your demographic information (duration: 1 – 3 minutes).

We would like to request that you be as open and honest as possible with your responses and avoid any perceptions of what you think a desired answer might be. The reliability of the data depends on your complete honesty for the sleep quality questionnaire. Please simply give the answer according to your opinion and your situation. Please try to make sure you have not inadvertently missed out any questions. For the reaction time task, please respond as quickly as possible.

Finally, we remind you that you are free to withdraw from the study at any point and if you feel uncomfortable answering any of the questions, you are free to not respond to those questions.

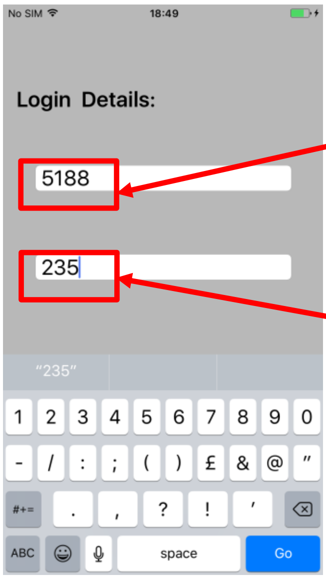
Thank you again for your participation.

ATW – Day 1: Session 2

1

Step 1:
Launch the **Reaction Time** app.

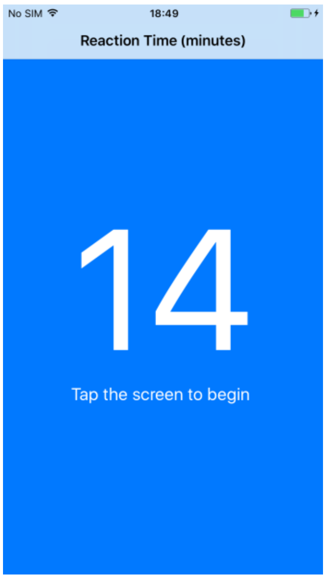
Step 2:
Enter both your **Volunteer Number** and **Password**, provided by the experimenter.

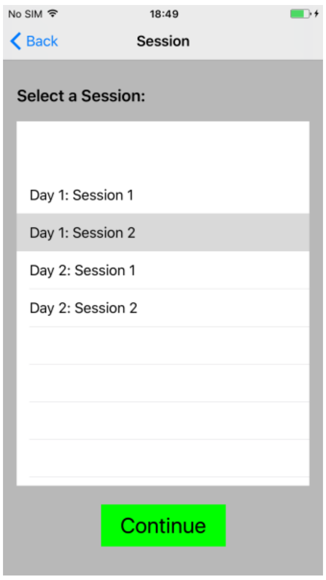


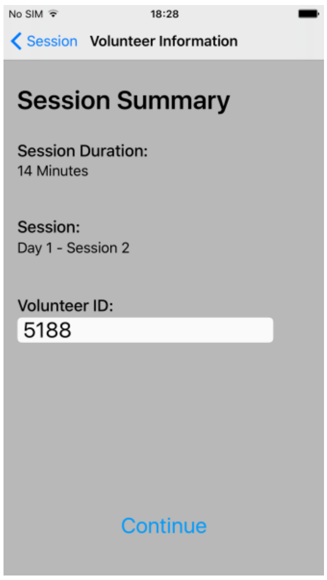
Enter:
Volunteer Number: 5188

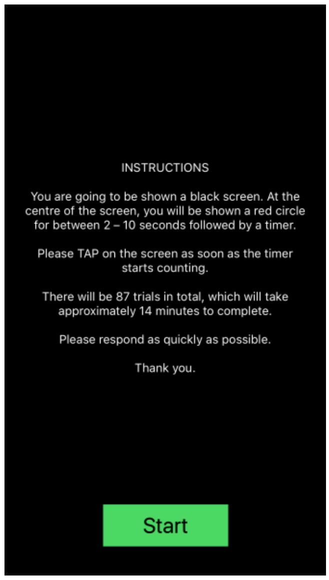
Enter:
Password: 504

Step 3:	**Step 4:**	**Step 5:**	**Step 6:**
Tap anywhere on the screen to continue.	Select **Day 1: Session 2**, then tap **Continue**.	Make sure your **Session** and **Volunteer ID** are correct, then tap **Continue**.	Follow the on-screen Instructions carefully and tap **Start** to begin.







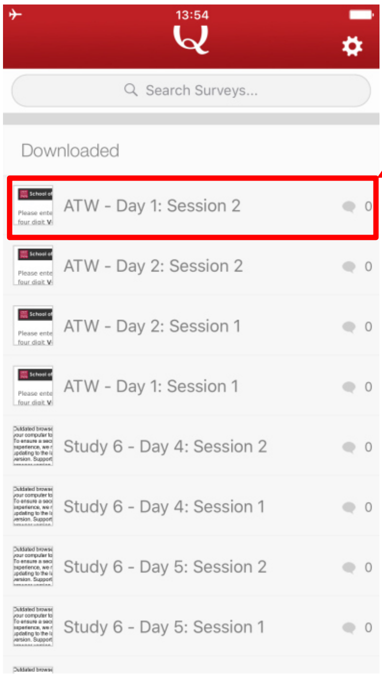


ATW – Day 1: Session 2

2

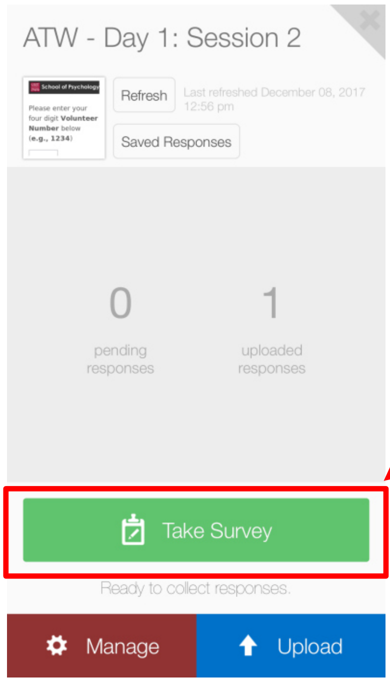
Step 1:
Launch the **Qualtrics Surveys** app.

Step 2:
Click on the **'ATW – Day 1: Session 2'** survey



Select:
ATW – Day 1: Session 2

Step 3:
Then tap on **'Take Survey'**



Tap:
Take Survey

Step 4:
Once you have completed the questionnaire, you are done!

Appendix VV

INFORMATION SHEET

LOW WORKLOAD BEFORE DIAGRAM

Thank you for agreeing to participate in this study. The aim of the study is to examine whether fatigue is associated with workload. This will be examined using the Psychomotor Vigilance Task, which will be administered using an Apple iPhone 6s Plus.

You will be kindly asked to complete a four-part study using a mobile device. These are divided into two different days, a LOW workload (i.e., spare driver) and HIGH workload (i.e., a difficult diagram), at two different time points (BEFORE diagram and AFTER diagram) using a mobile device. A total of £20 will be awarded to participants, upon successful completion of all four-parts. All payments will be made via pre-written cheques. Participation of this study will take approximately 15 – 20 minutes per session. All four-sessions combined will take approximately 60 – 80 minutes to complete.

This is the **LOW** workload day **BEFORE** diagram session. Participation of this session will involve completing:

1. A reaction time task (duration: 10 minutes);
2. A 12-item questionnaire on your sleep quality (duration: approximately 1 - 3 minutes).

We would like to request that you be as open and honest as possible with your responses and avoid any perceptions of what you think a desired answer might be. The reliability of the data depends on your complete honesty for the sleep quality questionnaire. Please simply give the answer according to your opinion and your situation. Please try to make sure you have not inadvertently missed out any questions. For the reaction time task, please respond as quickly as possible.

Finally, we remind you that you are free to withdraw from the study at any point and if you feel uncomfortable answering any of the questions, you are free to not respond to those questions.

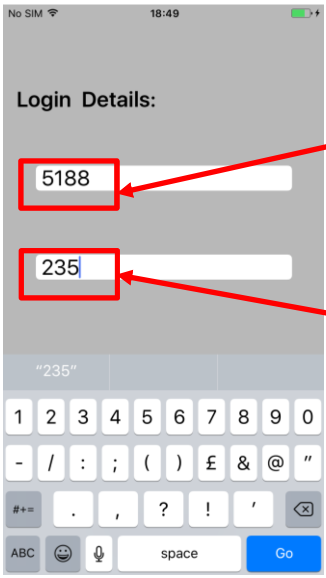
Thank you again for your participation.

ATW – Day 2: Session 1

1

Step 1:
Launch the **Reaction Time** app.

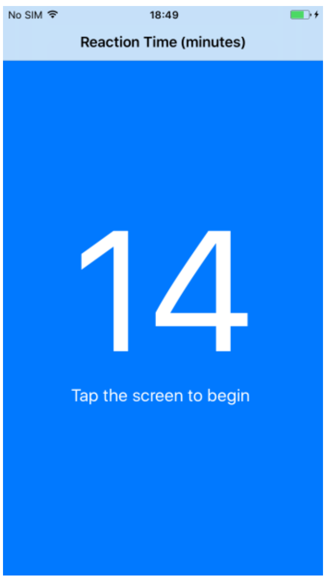
Step 2:
Enter both your **Volunteer Number** and **Password**, provided by the experimenter.

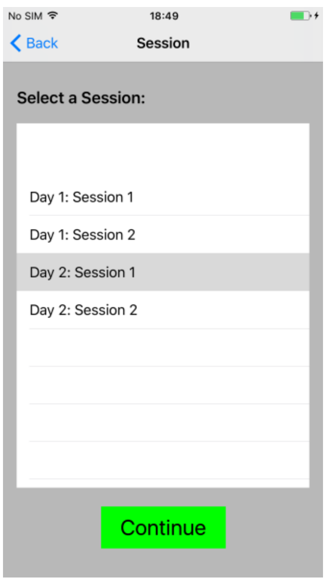


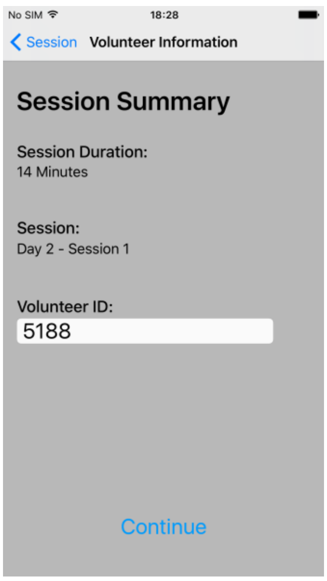
Enter:
Volunteer Number: 5188

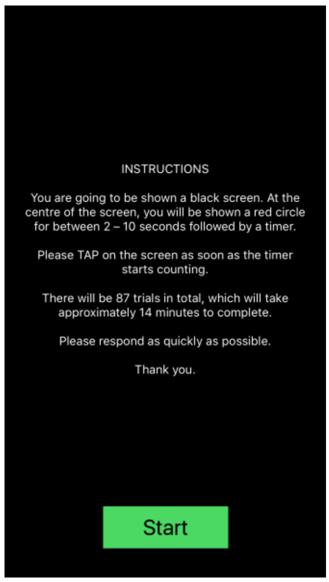
Enter:
Password: 504

Step 3:	**Step 4:**	**Step 5:**	**Step 6:**
Tap anywhere on the screen to continue.	Select **Day 2: Session 1**, then tap **Continue**.	Make sure your **Session** and **Volunteer ID** are correct, then tap **Continue**.	Follow the on-screen Instructions carefully and tap **Start** to begin.







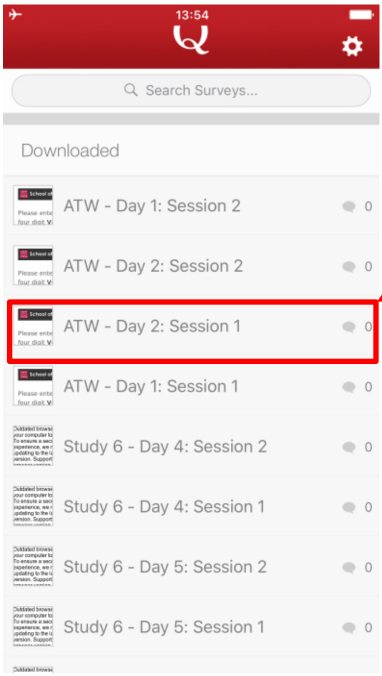


ATW – Day 2: Session 1

2

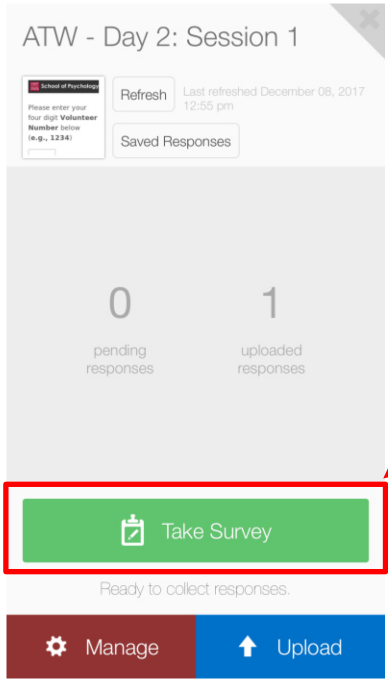
Step 1:
Launch the **Qualtrics Surveys** app.

Step 2:
Click on the **'ATW – Day 2: Session 1'** survey



Select:
ATW – Day 2: Session 1

Step 3:
Then tap on **'Take Survey'**



Tap:
Take Survey

Step 4:
Once you have completed the questionnaire, you are done!

Appendix YY

INFORMATION SHEET

LOW WORKLOAD AFTER DIAGRAM

Thank you for agreeing to participate in this study. The aim of the study is to examine whether fatigue is associated with workload. This will be examined using the Psychomotor Vigilance Task, which will be administered using an Apple iPhone 6s Plus.

You will be kindly asked to complete a four-part study using a mobile device. These are divided into two different days, a LOW workload (i.e., spare driver) and HIGH workload (i.e., a difficult diagram), at two different time points (BEFORE diagram and AFTER diagram) using a mobile device. A total of £20 will be awarded to participants, upon successful completion of all four-parts. All payments will be made via pre-written cheques. Participation of this study will take approximately 15 – 20 minutes per session. All four-sessions combined will take approximately 60 – 80 minutes to complete.

This is the **LOW** workload day **AFTER** diagram session. Participation of this session will involve completing:

1. A reaction time task (duration: 10 minutes);
2. An 8-item questionnaire on your workload (duration: approximately 1 minute).
3. A 10-item questionnaire on your demographic information (duration: 1 – 3 minutes).

We would like to request that you be as open and honest as possible with your responses and avoid any perceptions of what you think a desired answer might be. The reliability of the data depends on your complete honesty for the sleep quality questionnaire. Please simply give the answer according to your opinion and your situation. Please try to make sure you have not inadvertently missed out any questions. For the reaction time task, please respond as quickly as possible.

Finally, we remind you that you are free to withdraw from the study at any point and if you feel uncomfortable answering any of the questions, you are free to not respond to those questions.

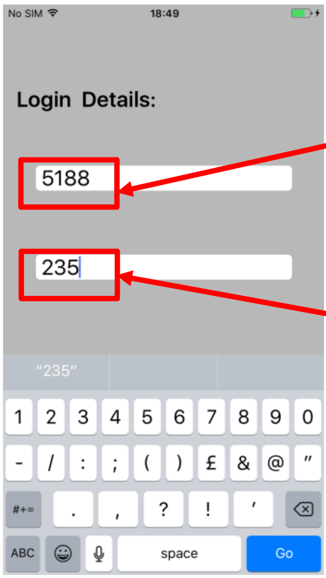
Thank you again for your participation.

ATW – Day 2: Session 2

1

Step 1:
Launch the **Reaction Time** app.

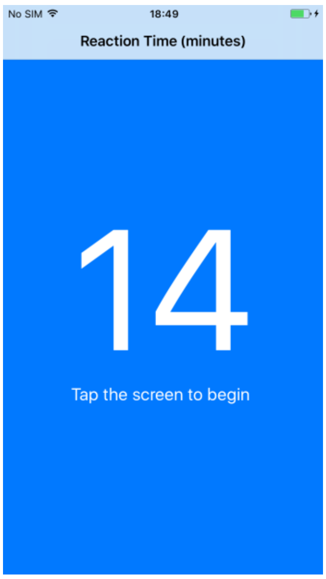
Step 2:
Enter both your **Volunteer Number** and **Password**, provided by the experimenter.

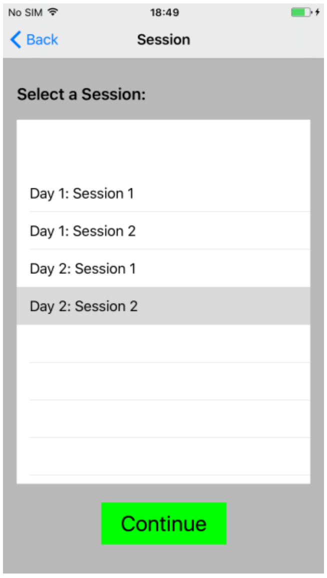


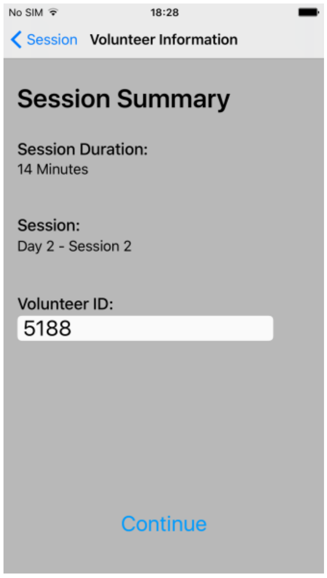
Enter:
Volunteer Number: 5188

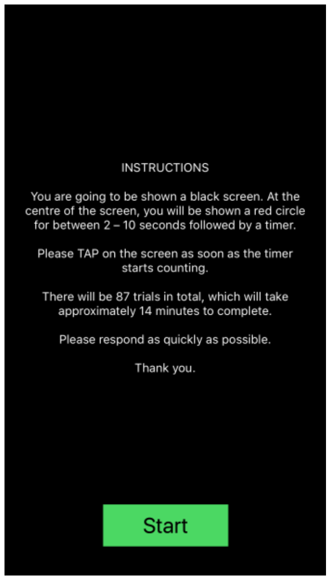
Enter:
Password: 504

Step 3:	Step 4:	Step 5:	Step 6:
Tap anywhere on the screen to continue.	Select **Day 2: Session 2**, then tap **Continue**.	Make sure your **Session** and **Volunteer ID** are correct, then tap **Continue**.	Follow the on-screen Instructions carefully and tap **Start** to begin.







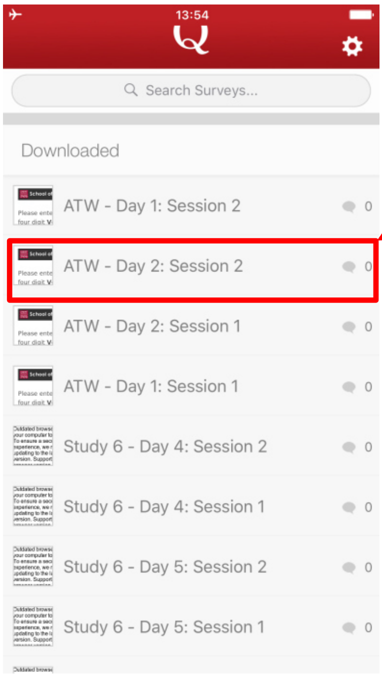


ATW – Day 2: Session 2

2

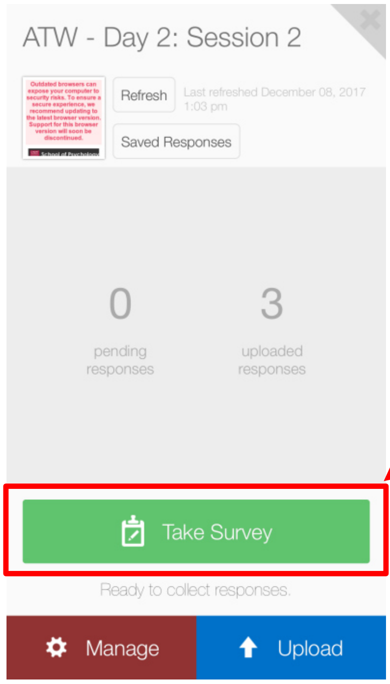
Step 1:
Launch the **Qualtrics Surveys** app.

Step 2:
Click on the **'ATW – Day 2: Session 2'** survey



Select:
ATW – Day 2: Session 2

Step 3:
Then tap on **'Take Survey'**



Tap:
Take Survey

Step 4:
Once you have completed the questionnaire, you are done!

Appendix BBB

DEBRIEF

Thank you for participating and completing all four-parts of the workload study, implemented on a mobile device. The aim of this study is to examine whether fatigue is associated with workload. In other words, whether cognitive levels of performance are associated with time of day.

Each participant's electronic consent to participate and corresponding data that you have provided will be held completely anonymous, and only intended for academic research, so that it is impossible to trace this information back to any individual. Only Professor Andy Smith and the researcher (Mr Michael Scott Evans) will have access to your data. This information will be stored and analysed for publication before it is destroyed, in accordance with the Data Protection Act 1998.

If you have any queries or concerns about the research, please contact either the researcher Michael Scott Evans, or the supervisor Professor Andy Smith by using the contact details below. If you are affected by any of the issues raised in the four online tasks, or from the sleep quality questionnaire, then there are a number of services available through the university that is able to offer support, which can be accessed using the following links.

Staff of Centre for Occupational and Health Psychology
Equality and diversity
http://www.cardiff.ac.uk/govrn/cocom/equalityanddiversity/index.html

Counselling service
http://www.cardiff.ac.uk/counselling/about/index.html

If you still remain unhappy with the support and wish to complain formally, you can do so via the Cardiff University Ethics Committee:

Secretary of the Ethics Committee
School of Psychology
Cardiff University
Tower Building
70 Park Place, Cardiff, Wales
United Kingdom, CF10 3AT
Tel: 029 2087 0360
Email: psychethics@cardiff.ac.uk

In addition, if you like to get any general information of the research findings such as published research articles and journals, please contact Michael Scott Evans (Doctoral candidate).

Thank you for your participation.

Michael Scott Evans
Doctoral Candidate
Centre for Occupational and Health Psychology
Cardiff School of Psychology
63 Park Place, Cardiff, Wales
United Kingdom, CF10 3AS
Email: EvansMS3@cardiff.ac.uk

Professor Andy Smith
Director/Supervisor
Centre for Occupational and Health Psychology
Cardiff School of Psychology
63 Park Place, Cardiff, Wales
United Kingdom, CF10 3AS
Email: SmithAP@cardiff.ac.uk