# Performance Shaping Factors Affecting Driver Safety-Related Behaviour in Urban Rail Systems: Tyne & Wear Metro Case

**Aleksandrs Rjabovs**

This thesis is presented for the degree of

Doctor of Philosophy

School of Mechanical and Systems Engineering

Newcastle University

June 2017

# Abstract

It is accepted that train drivers' safety performance is affected by numerous performance shaping factors (PSF). Design of the physical environment is among these factors. Even though the body of knowledge in rail human factors is increasing, it is limited as it is often i) reactive, ii) focusing mainly on single type incidents, iii) prioritising high profile accidents, iv) not always fully addressing existing risk profiles. Railway systems with different design features are usually grouped together for research purposes thus disregarding the fact that system design can alter effects of the PSFs. This is especially true for urban rail systems. A combination of concurrent and sequential research in this mixed methods thesis has investigated PSFs associated with metro systems design, using the Tyne & Wear Metro system as its application case. The PSFs embedded in everyday operations have been studied on different system levels through historic incident analysis, drivers' surveys, semi-structured interviews, eye-tracking and simulation experiments. Some of the established methodologies have been adapted in order to address the research objectives set. Novel approaches have been developed for the deployment of in-service eye-tracking using dynamic areas of interest and the development of a low-cost high fidelity simulator using gaming software and hardware. Selected station layouts have been assessed through measures of workload, stress and signal checking behaviour thus supporting PSF inter-dependence. The results suggest the influence on the performance of arrival and departure procedures of the angle between a signal, a driver and a mirror. Among the latent conditions potentially inducing incident propagation are passenger levels, the platform side, informativeness of design elements, openness and lighting conditions of a station, and distances from a stopping position to other elements of the station design.

# Dedication

This thesis is dedicated to my parents who supported and encouraged me all my life to aspire for something bigger. Their hard work and commitment to a goal has been the best example I could ever ask for.

# Acknowledgements

I would like to thank my supervisors, Dr Roberto Palacin and Professor Mark Robinson, for their encouragement, support and patience in this long journey. Their contribution helped me to power through difficult times and pace myself when I was getting too enthusiastic. I am also forever grateful for being a part of NewRail – Centre for Railway Research. The amazing people I met there and the way they inspired me every day will always be in my heart.

This thesis would not have happened without enormous support from DB Regio Tyne & Wear and Nexus. I am truly grateful to Ken Blyth, John Naisby and John Kear for their valuable time and hours of interesting discussions, Darren Armstrong for insights into metro driver's job, Peter Stout for continuous encouragement and all of Tyne & Wear Metro employees at South Gosforth Control Centre for their invaluable help.

I am highly indebted to Dr Joan Harvey and Jonathan Doughty for their assistance with the questionnaire study.  I would also like to express thanks to all of the Tyne & Wear Metro drivers who spared their free time to participate in the experiments and were some of the most forthcoming and accommodating people I have ever met.

This research has been supported by the Institute for Sustainability at Newcastle University, through the Sir James Knott and Ridley PhD Scholarship, for what I am very grateful.

Finally, I would like to acknowledge the big input my wife and close friends made to this thesis through encouragement, motivation and helping me to forget about the work at times when I needed this.

# Table of contents

# List of tables

# List of figures

# Glossary

| Abbreviation | Stands for |
|---|---|
| **AFD** | Average Fixation Duration |
| **AOI** | Area of Interest |
| **ATO** | Automatic Train Operation |
| **ATP** | Automatic Train Protection |
| **ATS** | Activation-Trigger-Schema |
| **AWS** | Automatic Warning System |
| **CCTV** | Closed Circuit Television |
| **CSDE** | Correct Side Door Enable |
| **DAS** | Driver Advisory Systems |
| **DMI** | Driver-Machine Interface |
| **DOO** | Driver Only Operation |
| **FOW** | Field of View |
| **GPL** | Ground Position Lights |
| **HF** | Human Factors |
| **HFACS** | Human Factors Analysis and Classification System |
| **HRA** | Human Reliability Analysis |
| **HVAC** | Heating, Ventilation and Air Conditioning |
| **LED** | Light-Emitting Diode |
| **LRA** | Low Rail Adhesion |
| **NR** | Network Rail |
| **NTS** | Non-Technical Skills |
| **OTMR** | On-Train Monitoring Recorded |
| **PPMC** | Pearson Product-Moment Correlation |
| **PSF** | Performance Shaping Factor |
| **PTI** | Platform Train Interface |
| **RAIB** | Rail Accident Investigation Branch |
| **RSSB** | Rail Standards and Safety Board |
| **SA** | Situation awareness |
| **SaSSPaD** | Start against Signal Pass at Danger |
| **SDC** | System Desk Controller |
| **SDC** | System Desk Controllers |
| **SPaD** | Signal Pass at Danger |

| | |
|---|---|
| **SPSS** | Statistical Package for Social Sciences (developed by IBM) |
| **SRC** | Spearman Rank Correlation |
| **SUS** | System Usability Scale |
| **T&W** | Tyne & Wear |
| **TASC** | Train Automatic Stop Control |
| **TFC** | Total Fixation Count |
| **TFD** | Total Fixation Duration |
| **TOC** | Train Operating Company |
| **TPWS** | Train Protection and Warning System |
| **TRACEr** | Technique for the Retrospective and Predictive Analysis of Cognitive Errors |
| **UTO** | Unattended Train Operations |

# Chapter 1. Introduction

The purpose of this thesis is to investigate effects of performance shaping factors (PSFs) associated with design aspects of urban rail systems and their influence on safety related behaviour of metro train drivers. Tyne & Wear Metro is used as a case study to establish the relevant PSFs, their inter-dependency and how they differ from those applied to mainline railways.

## 1.1 Background

Railway systems are one of the safest modes of transport, but they are not risk-free (European Transport Safety Council, 2003; RSSB, 2016); incidents do occur, in a range of magnitude, and sometimes with severe consequences. Despite being the safety leader in European railway industry, hundreds of major passenger and workforce accidents happen in the United Kingdom annually (RSSB, 2016). The majority of incidents lead to passenger minor injuries, but fatalities also happen. Nevertheless, there were no train accidents recorded with on-board fatalities between 2006 and 2016 (RSSB, 2016). This has been achieved with increased attention to collision-related issues. Among those are crashworthiness, track quality and Signals Passed at Danger (SPaD). The drive to reduce collisions and subsequently SPaDs has its origin in most high-profile rail accidents in the UK in the past 25 years being a result of a driver passing on a red aspect, e.g. Ladbroke Grove collision in 1999. Since then the SPaD has become an important unit in assessment of driver-related safety performance of train operating companies (TOC) and railways. There were still 277 SPaDs in 2015/16 in the UK despite introduction of train protection and warning system (TPWS), improved SPaD reporting and analysis, and a wide range of other SPaD-focused initiatives (RSSB, 2016).

Railway safety research is rather reactive and is often initiated in response to a high profile incident. Numerous casualties and injuries are the main pre-requisites for a “high-profile accident” status. However, the UK statistics shows that cumulative risk to passengers is higher from “individual accidents”. Masaaki *et al.* (2015) claim that SPaDs and overspeeding are the two most common incidents resulting from human error, but Rail Standards and Safety Board (RSSB, 2016) demonstrate that risk from platform-train interface (PTI) and on board accidents is more than five times higher than from train accidents. Moreover, the PTI accidents have the highest risk of fatalities among passengers. Such PTI accidents rarely involve more than one passenger so might go unnoticed, especially in smaller networks. Even though there has been an increased research focus on PTI accidents, the existing body of knowledge is rather fragmented. Incidents such as wrong-side door activations rarely become main research focus. On the other hand, SPADs are well-covered in the literature (Multer *et al.*, 2015). Moreover, “*SPADs continue to be high priority incidents for the GB rail industry*” (p. I) (Gibson, 2016). The reactive approach along with an increased focus on prevention of “high-profile accidents” led to a lack of synergy between research areas and the actual risks in the industry.

Despite a rich body of knowledge on SPADs, it is not easily transferable across countries and requires adaptation nationally, e.g. in Australia and New Zealand (Naweed and Rainbird, 2014). Moors *et al.* (2015) advocated that the existing research, if applied to a new system, has to consider individual characteristics of the system, e.g. demand on human performance. British railway industry consists of wide variety of different systems, including urban rail, mainline passenger and freight services, heritage lines. However, most of the UK railway-related research is driven by mainline operators through organisations such as RSSB and ATOC (Association of Train Operating Companies). This creates a situation of “one size fits all” where tools developed on mainline railways are applied to metro railways, e.g. various SPAD checklists (RSSB, 2006) are used in Tyne & Wear Metro, or conclusions are drawn from generalisation across different systems (RSSB, 2005). Such lack of distinction between different systems means that tools and solutions offered are not exactly tailored and cannot achieve maximum efficiency, if any. In terms of urban rail, the main issue is disregard of considerable differences of such systems and subsequent rareness of metro-specific research.

Even with larger number of signals encountered, collision risks are rather low in metro systems due to usual presence of automatic train protection (ATP) systems. These systems monitor drivers’ speeds and moving authority at certain locations. If violation is spotted, the system applies brakes automatically. With lighter trains and lower maximum speeds, the trains can be stopped in very short distances. For example, Indusi ATP system used in Tyne & Wear Metro can stop a 2-car train from 80 km/h in less than 150 metres compared to 2 km for some mainline trains. However, significant jerk experienced by passengers in such situations creates additional risks in terms of major injuries (Powell and Palacín, 2015). The other important factor in SPaD propagation is drivers’ route knowledge (Naweed, 2013; Gibson, 2016). Metro drivers operate in closed systems which are more uniform.

They do not have to deal with complicated gantry arrangements, variety of rolling stock and perplexing junction layouts. Hence metro drivers should be better than mainline train drivers in this parameter. The event rate of task and route knowledge elements in the urban rail environment is very high, which significantly affects strategies employed by drivers (Naweed and Balakrishnan, 2014). Higher concentration of station stops further induces risks associated with the PTI. In mainline railways, the combination of 4-aspect signalling, longer stopping distances and limited signal sighting requires predictive, feed-forward control (Stanton and Walker, 2011). It is not the case for the metro systems where simpler 2- or 3-aspect signalling is used for capacity reasons. Hence simple “see red aspect – apply brakes” models are more often encountered in urban railways.

All of the above factors combined create a different risk profile to that found in mainline railways. Even though signal sighting, PTI risks, route knowledge and other aspects of train driving are explored in the literature, it is usually done in the context of mainline rather than urban rail. Moreover, the literature which investigates these issues from system design point of view started to appear only recently.

This thesis approaches the issues of drivers’ safety related performance in urban railways through the case of Tyne & Wear Metro. This system has features typical to both metro and suburban rail systems, including underground running at own infrastructure, mixed traffic on Network Rail infrastructure, Driver Only Operations (DOO). Many of these features can be encountered in urban rail systems across the world. Hence the findings can be extrapolated to those systems.

## 1.2 Railway incident propagation

Modern railways still depend heavily on human operators who are known to make errors. Scientists acknowledge a state of zero errors is often non-achievable (Kontogiannis and Malakis, 2009) and human operators are often solely blamed for incidents (Mackieh and Cilingir, 1998; Stanton and Salmon, 2009). However, there is a move to more in-depth investigation of human error through exploration of causal factors (Plant and Stanton, 2012; Madigan *et al.*, 2016). Incidents are caused by a combination of technical, systematic and human factors (Woods *et al.*, 2010). Alignment of many factors, which are not sufficient to cause an incident individually, can be required for an incident to happen (Bogner, 2002).

The famous “Swiss Cheese Model” introduced by Reason (1997) contains an important notion of human errors being a result of latent factors embedded in a system. It identifies workplace and organisational factors as both important contributors to unsafe acts and latent conditions in failures of system defences (Figure 1). The critics of the model claim that it is being too theoretical and does not take into account relationships between the planes and causal factors (Shappel and Wiegmann, 2000; Reason *et al.*, 2006). However, it represents a simple choice faced by the safety-critical industries between adding more defences or addressing the latent conditions.

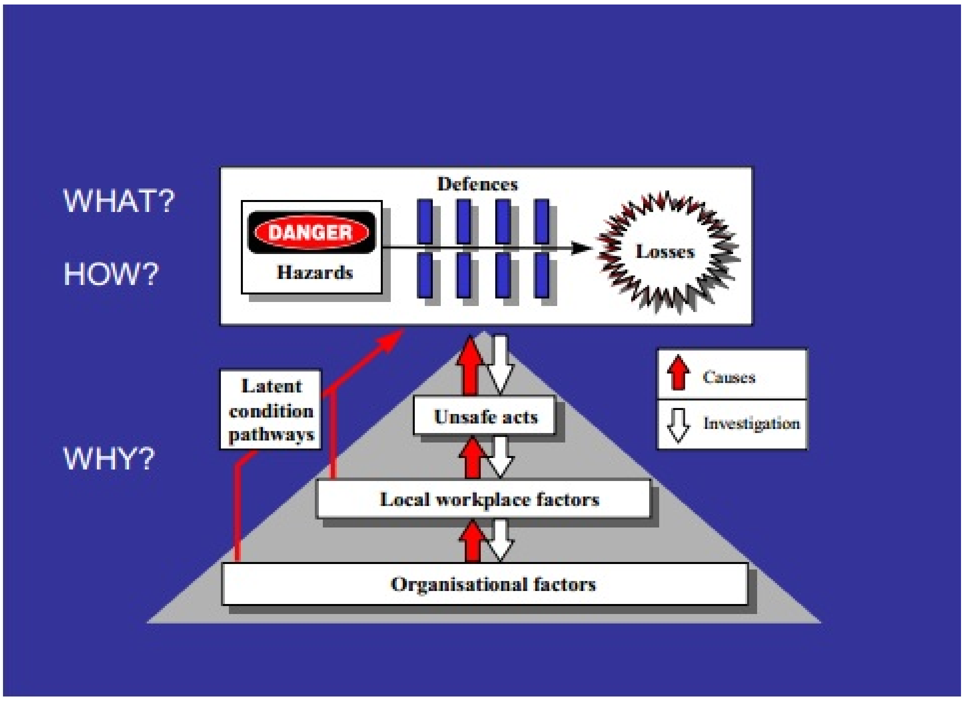


**Figure 1. Reason’s accident causation model, version of 1997 (Reason, 1997)**

## 1.3 Rail human factors

The railway industry uses automation as the main way of increasing system defences. It can eliminate a lot of latent conditions but also create many new causal factors. Automation is a costly activity, which becomes even more expensive if retro-fitting is considered. Only 166 km of total 2335 km of metro lines in the world were used for unattended train operations in 2009 (The European Rail Research Advisory Council, 2009). Semi-automation through various driver advisory systems is becoming popular but possesses similar challenges to full scale automation. Moreover, human-centred design is very important in order to actually introduce positive effects not create more negative ones. These technologies support the downward trends in incident propagation but do not positively influence drivers’ performance. Research suggests that train drivers’ safety-related performance has not improved at all between 2004 and 2013 considering only operations with technical solutions disabled (Russell *et al.*, 2013). In other words, drivers keep making the same amount of mistakes but novel technical solutions prevent some errors from becoming incidents.

De Egea *et al.* (2013) found out that approximately 80% of the risk in the railway industry still can be attributed to front line staff, whereas most investment is streamed into the technical domain. Human Factors (HF) is considered a suitable approach to address all aspects of safety, in safety-critical industries and systems (Vogt *et al.*, 2010). Rail human factors discipline had a slower start than HF research in other safety-critical industries, but has gained momentum in 21st century (Wilson and Norris, 2006; Zeilstra and Van Der Weide, 2013). The growing relevance of this discipline is evidenced by the increased involvement of human factor specialists in the design of railway systems (Crawford *et al.*, 2013).

Human Reliability Assessments (HRA) extensively use Performance Shaping Factors (PSFs) in order to display quantified human factors' influence on human performance. PSFs absorb the organisation and management influences on human error probability and are specifically selected for each system (Belamy and Geyer, 2007). Gibson *et al.* (2009) define PSFs as "*A state which cannot in itself be defined as an error but increased the likelihood of one or more errors. They are non-optimal states which are accepted features of a task, management system or design*" (p. 13). This description does not include acknowledgement of potential positive effects of these factors. Along with increased use of systematic approach to incident investigation (Ryan *et al.*, 2010; Plant and Stanton, 2012), performance shaping factors become an important concept in understanding incident causation.

Due to previously discussed differences of mainline and metro systems, the PSFs associated with system design might require more adaptation from mainline railways compared to organisational, team and personal factors. Different rail PSF taxonomies (Kyriakidis *et al.*, 2012b; Gibson *et al.*, 2013b; Gibson, 2016) include factors associated with design of physical environment.

Gourlay *et al.* (2013) highlighted the importance of system design related factors to drivers' performance. Gibson (2016) in his assessment of Incident Factors (RSSB classification system for causal mechanisms) emphasise equipment and work environment design as PSFs. Factors associated with physical and technological environment were found present in approximately a quarter of rail incidents (Madigan *et al.*, 2016). Naumann *et al.* (2016) advocate that system design of drivers' workplace is a central influencing factor in drivers' performance. Elliott (2010) lists areas of particular importance to the rail HF, including anatomy and anthropometry, physiology and physiological workload.

One of the most significant characteristics of using PSFs, is the recognition that human error is caused by a mix of different factors, thus acknowledging inter-dependence (De Ambroggi and Trucco, 2011). However, the complex inter-dependence of the PSFs is yet to be understood (Basacik *et al.*, 2015; Naumann *et al.*, 2016). Discerning the PSFs can help create safety-critical systems that include a human operator as an asset, rather than a risk carrier. It also provides great opportunities for system engineers who potentially could "design out" or contain adverse PSFs related to the physical environment.

Some of the work done in this area includes re-design of a mainline train cab to reduce harmful low-frequency vibrations (Johnson *et al.*, 2009), relationship between infrastructure features and train drivers' arousal levels (Keun Sang and Ohkubo, 1994; Yang *et al.*, 2012). It is recognised that elements of immediate physical environment, e.g. driver-machine interface (DMI) or a driver desk, are closely connected with other negative PSFs such as workload and poor situational awareness (RSSB, 2004a; Blanchard, 2013; Hitchcock *et al.*, 2013; Kecklund *et al.*, 2013; Sumpor *et al.*, 2013; Van Der Weide *et al.*, 2013). Outside the driver's cab signal designs, and how those contribute to signal sighting, have been researched (Human Engineering Ltd, 2006; Li *et al.*, 2006; RSSB, 2007b; Elliott, 2012). Environmental factors are of great importance to train drivers' performance too with suggestions that effective system design can mitigate those (RSSB, 2008; 2012).

The literature review (Chapter 2) will expand on many of the above mentioned research and how it mostly focuses on mainline railways. The current body of knowledge approaches a research problem from task design point of view. For example, signal sighting studies do not consider how task demands and priorities change with variation in operational environment or design features. Adding an additional dispatch tasks in DOO systems might significantly affect drivers' interaction with station signals. In a sense, many studies do not investigate PSFs with inter-dependencies considered.

## 1.4 Rationale

The existing body of knowledge clearly identifies risks embedded in railways but often fails to address those in urban rail context. However, it can be expected that many of the findings from mainline railway are applicable to metro systems but effect on other PSFs needs to be understood. Such a combination of known risks and potentially transferable knowledge creates an opportunity for a leap forward in understanding of urban rail drivers' performance. Moreover, with many urban rail design and operational features potentially transferable to mainline railways, an importance of such research is even higher.

With growing urbanisation and demand for travel among city dwellers, urban railways simply cannot afford costs associated with incidents even if safety implications are minor (Madigan *et al.*, 2016). The aspiration for shorter headways and better reliability of the system means that tolerance of disruptions will continue to decrease. According to Naweed and Rainbird (2014), the incidents stemming from human errors advocate use of human factors. Previous research in HF and PSF demonstrates that performance improvements can be achieved in a cost-effective manner. The cost of retrofitting and design alterations leads to situations when system engineers and designers have only one attempt to make it right. In the context of this thesis, “making it right” means “engineering out” as many adverse PSFs as possible through thorough design considerations.

System design PSFs are often considered from SPaD point of view thus only meaning increased focus on signal design and sighting research. However, the risk profile demonstrates importance of the PTI related hazards and addressing those. When research focuses on certain tasks of train drivers, e.g. signal sighting, this can lead to a gap in more holistic understanding of influence of a wider physical environment on safety-related performance of front line staff. This thesis investigates system design PSFs through a series of experiments with Tyne & Wear Metro. However, if successful, the methodology should be applicable to other urban rail systems providing operational and design differences are taken into account.

## 1.5 Objectives

With the above rationale in mind, the thesis addresses the following research questions:

1. Is the existing mainline research applicable to urban rail systems?
2. What design features act as performance shaping factors in metro driver related incident propagation?
3. What design and procedure modifications to existing urban rail systems can mitigate the effects of these PSFs?

In order to address the research questions, the following objectives are set:

A. To explore the current body of knowledge on infrastructure design PSFs and how it applies to Tyne & Wear Metro;
B. To identify design features associated with changes in drivers' performance objectively through analysis of the Tyne & Wear Metro past incident data;
C. To investigate the identified design features subjectively from drivers' point of view and in context of their everyday operations;
D. To carry out an in-field assessment of the PSF influences on drivers' performance;
E. To propose a set of measurable actions for mitigations of the identified adverse factors;
F. To assess the effectiveness of the proposed mitigations in the controlled environment.

Objective A is related to research question 1. Objectives B and C address both research questions 2 and 3. Finally, research question 3 is further explored through objectives D, E and F.

## 1.6 Thesis structure

The thesis consists of the nine chapters as follows:

Chapter 2 presents an overview of the existing body of knowledge on system design PSFs in railway industry. It expands on the introduction statements about necessity to investigate urban rail systems in depth.

Chapter 3 defines research strategy and research design of the thesis through an argument about the most applicable research philosophy. The research tools and methods are presented. The chapter describes how those fit into the selected research design.

Chapter 4 describes the design and operational features of Tyne & Wear Metro. A categorisation of station layouts and examples of design deviations are defined.

Chapter 5 analyses Tyne & Wear Metro historic incident data to identify design features potentially affecting drivers' performance. The objective findings of frequency distribution and correlation analyses are enriched with subjective results from the semi-structured interviews and workshops. This chapter addresses objectives 2 and 3.

Chapter 6 presents a collection of Tyne & Wear Metro drivers' perceptions and opinions on the influence of the selected design features on their safety related performance (objective 3). Both subjective and objective data is collected and analysed from the mixed methods questionnaire study. The chapter discusses factors influencing drivers' opinions and assessments of the PSFs.

Chapter 7 identifies four elements of the system design which can be critical to drivers' performance. The designs of those elements are assessed as PSFs during the in-field cognitive data collection (objective 4). Measures of drivers' workload and stress are used to explore the differences in designs of each element. The chapter also suggests a set of measurable actions that can influence how effective the investigated PSFs are (objective 5).

Chapter 8 continues investigation of the designs of the four elements in a more controlled environment. The experimental design allows testing the measures suggested in the previous chapter. The findings provide assessment of efficiency of the suggested measures.

Chapter 9 concludes the thesis and summarises the most important findings. Contributions to knowledge and a scope for further research in the area are outlined.

## 1.7 Contribution to the body of knowledge

The thesis contributes to the body of knowledge as follows:

- Evidence-based identification of system design PSFs in the context of urban rail systems;
- Validated set of design mitigations for the above PSFs;
- Validated low-cost methodology for assessment of the PSFs and mitigations measures.

The work described in the thesis contributed to several peer-reviewed international conference and journal publications as follows:

- Rjabovs A., Palacin R., Robinson M. (2014). ‘Cab and system design influence on metro drivers' performance: preliminary study’. Transport Research Arena; April 2014; Paris.
- Rjabovs, A. and Palacin, R. (2015) ‘Attitudes of Metro Drivers Towards Design of Immediate Physical Environment and System Layout’, *Urban Rail Transit,* 1 (2), p. 104-111.
- Rjabovs, A. and Palacin, R. (2016) 'The influence of system design-related factors on the safety performance of metro drivers', *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit,* (ePub ahead of Print).
- Rjabovs, A. and Palacin, R. (2016) ‘Design and layout of the physical environment in a metro system: appraisal of Tyne and Wear Metro drivers’ perceptions’. WCTRS, July 2016; Shanghai;

# Chapter 2. Literature review

This chapter introduces the issue of human performance with a focus on railway domain and system design as a performance shaping factor. It starts with an explanation of a wider human error concept and how it applies to railway industry. Then current academic advancements in terms of system design PSFs are reviewed. The trends in academic research and gaps in the body of knowledge are identified, particularly scarce literature on urban rail systems and PSFs in everyday DOO operations.

## 2.1 Human error

Each driver-related incident includes some sort of a human error. Even though the thesis approaches these errors from a systematic point of view, it is necessary to understand the failure mechanisms involved. The modern concepts of human errors claim that there is rather limited number of such mechanisms. Without understanding these concepts, investigating causal factors is significantly more complicated as both topics are inter-connected.  Front-line staff errors can generally be divided into two categories. As train driving depends on drivers' actions, the categories are based on performing wrong actions or not performing correct actions.

### *2.1.1 Human actions*

With human actions being a centrepiece of human performance, it is important to understand those first.  Norman (1981) outlined a theory of actions that use schemas. The modern schema concept was introduced by Neisser (1976), but the schema theory can be tracked as far as Ancient Greece (Plant and Stanton, 2012). The theory suggests that every action sequence is controlled by schemas – organised memory units or sensori-motor knowledge structures (Norman, 1981). Norman's activation – trigger – schema (ATS) system is based on activation and selection of these structures where specific conditions are required for triggering. The structures are also divided into parent and child ones. Parent schemas correspond to intention but a child schema – to a part of an action sequence. In this concept, even the simplest action sequences consist of a combination of different level knowledge structures and an error can occur at any of the ATS stages. For example, an incorrect schema can be activated or a schema is triggered by a wrong set of conditions.

Rasmussen (1983) offered his view on the process by presenting the skill-rule-knowledge framework. In comparison to the ATS system, the framework is mainly focused on skilled human operators of industrial installations. It consists of three levels – skill-based, rule-based and knowledge-based. Figure 2 provides a visualisation of the framework.

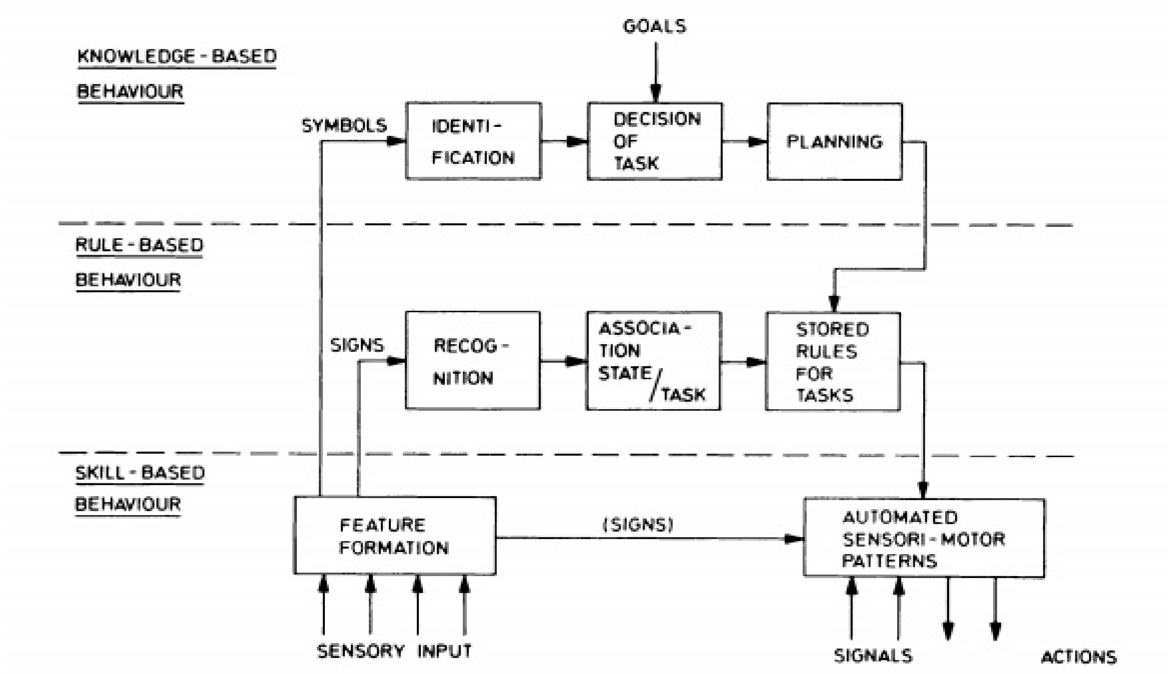


**Figure 2. Simplified illustration of the Rasmussen's (1983) skill-rule-knowledge framework**

The skill-based level is responsible for actions in response to a statement of intention. Actions are based on stored patterns of pre-programmed instructions and occur without conscious control. The rule-based level is necessary for solving familiar problems with actions being controlled by stored rules. Finally, the knowledge-based level is responsible for actions in unexpected or novel situations when analytical skills and knowledge base are required. Rasmussen's framework also appreciates multi-dimensionality of complex tasks, e.g. train driving, which can be carried on all 3 levels simultaneously. It is similar to the ATS concept but providing more specific explanation for different levels of a task.  In terms of train driving, a lot of it happens on the rule-based level. The knowledge-based behaviour, according to Rasmussen (1983), is mostly applicable to unfamiliar situations where a new plan has to be developed based on environment assessment and person's goals.

Information perception is also different between the levels of the framework. For example, observing a red signal on the rule based level will mean activation of braking sequences. The same visual information on the knowledge based level can trigger considerations of potential reasons for this signal, development of a forecast for a system state ahead and aligning it with, for example, punctuality goals.

However, it is a complicated process involving actions on many levels. Work by Reason (1990) shows that, in conditions of under-specification, a human cognitive system is inclined to choose responses that are most frequent in similar situations. Hence if a train driver never before encountered a red signal and never learnt about red signals meaning stop, he or she might treat it as a green signal due to missing information being substituted with stereotypical assignments for signal sightings.

### *2.1.2 Erroneous actions*

Authors of the human performance models also investigate what can hinder this performance. According to Neisser (1976) and Norman (1981), an erroneous action occurs when a defective schemata is selected or a faulty activation occurs. Typically, several types of erroneous actions are distinguished. For example, Norman (1981) describes three different categories with many sub-categories based on slips (Table 1). It is important to note, that the future research seem to substitute the term “slip” with “error” when quoting Norman’s work (Stanton and Walker, 2011). In fact, such substitution helps understanding this classification better as the more recent work specified slips as only one of several error types.

| Error Categories | Error Source |
|---|---|
| **Slips that result from errors in the formation of the intention** | Errors in determination of goals, decision making and problem solving |
| | Mode errors |
| | Description errors |
| **Slips as a result of faulty activation of schema** | Unintentional activation |
| | Loss of activation |
| **Slips as a result of from faulty triggering of schemas** | Failure to trigger |
| | False triggering |

**Table 1. Classification of errors adapted from Norman (1981)**

Reason (1990) offered his human error classification with 4 categories included. As it can be seen from Table 2, Reason’s classification relates these categories to planning and execution. Slips and lapses happen when a correct plan is not executed due to attention or memory failures. Whereas mistakes occur due to a plan not being correct in the first place. The classification also includes intentional (routine) and unintentional (exceptional) violations. In his later work, Reason suggests dividing mistake into failures of expertise (rule-based level) and lack of expertise (knowledge-based level) (Reason, 1997). Reason (1990) also categorises failure modes (error types) by performance levels (Figure 3), which are similar to the skill-rule-knowledge framework by Rasmussen (1983).

| **Basic error types** | | | |
|---|---|---|---|
| Slip (attention failure) | Lapse (memory failure) | Mistake (planning) | Violations |
| **Examples of error types** | | | |
| Misperception, action intrusion, omission of action, reversal of action, misordering of action and mistiming of action | Omitting of planned actions, losing place in action sequence, forgetting intended actions | Misapplication of good procedure, application of bad procedure, poor decision making, overconfidence, failure to consider alternatives | Intentional and unintentional |

**Table 2 Basic error types with examples (adapted from Reason, 1990). Source: Stanton and Walker, 2011**

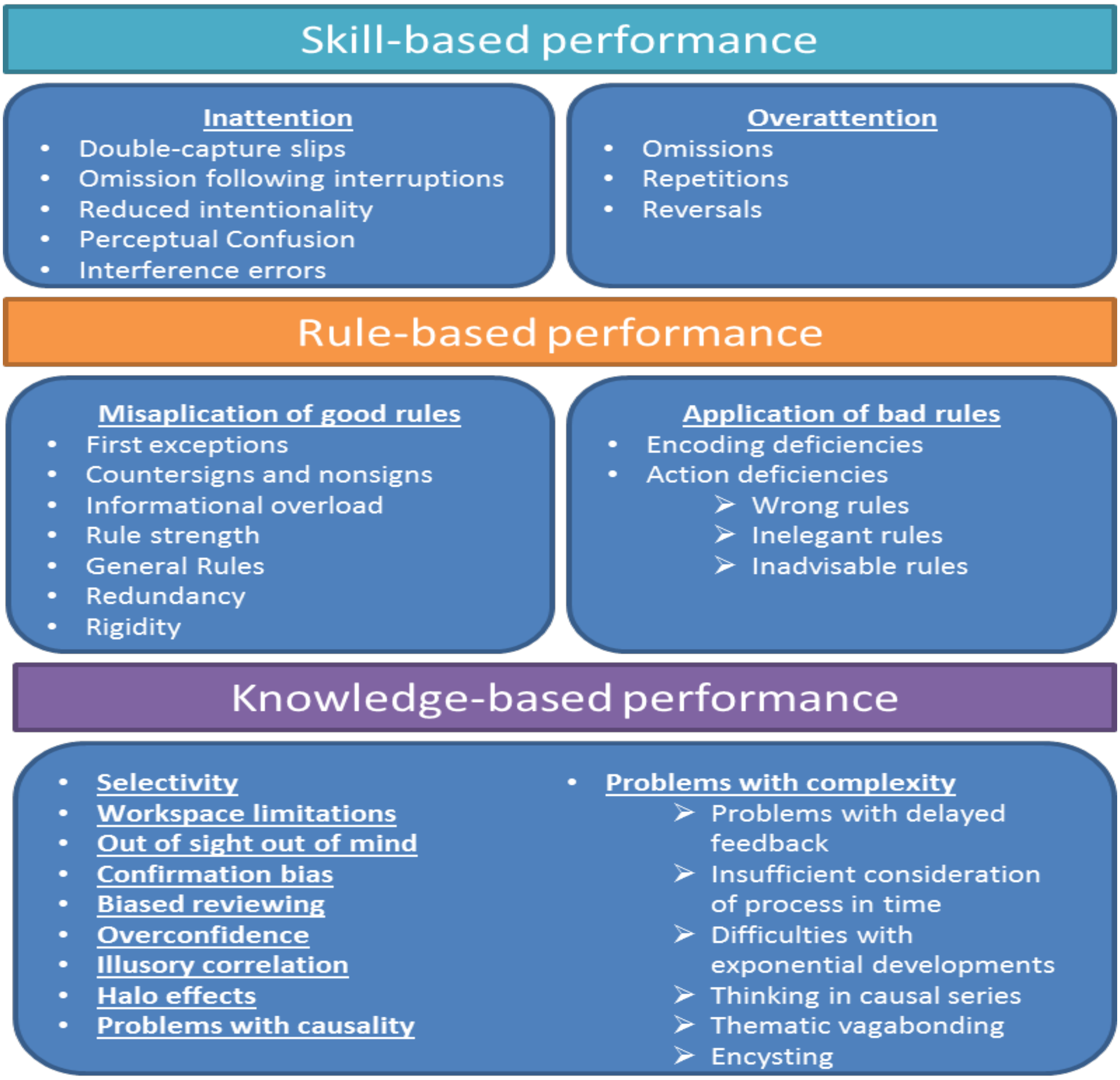


**Figure 3. Failure modes at different performance levels (adapted from Reason, 1990)**

The Reason's and Norman's models provide general error definition which is supposed to fit all situations. However, an increased interest in industry-specific understanding of the human error prompted development of many purpose built taxonomies across various domains.

### 2.1.3 Human error categorisation in safety-critical systems

Reason's human error model has been applied to various industries. For example, Human Factors Analysis and Classification System (HFACS) (Shappel and Wiegmann, 2000), developed for the US Federal Aviation Administration, uses Reason's human error concepts. The HFACS  has been successfully applied to other safety-critical industries such as railways (Baysari *et al.*, 2008) and shipping (Celik and Cebi, 2009). Zhang *et al.* (2004) developed their taxonomy of medical errors based on Reason's definition of human errors and Norman's action theory (Norman, 1988). The importance of an industry-focused taxonomy was demonstrated by Reason *et al.* (1990) who applied his own model to car driving. The resultant categorisation has only one lapse but many mistakes and slips. However, this proportion can be different in other industries. As shown by Stanton and Salmon (2009), understanding human errors is the first step to grasping causal factors and potential mitigations. Hence dedicated human error taxonomies allow building customised models of causal factors.

No universally agreed taxonomy of human errors in the railway industry exists. RSSB developed human error identification tool called '*Technique for the Retrospective and Predictive Analysis of Cognitive Errors* (TRACEr-rail)' (Baysari *et al.*, 2009). The TRACEr-rail framework, adapted from air traffic control domain, consists of eight different taxonomies on task errors, cognitive domains, and error recovery and detection. In their comparison between TRACEr-rail and HFACS, Gibson *et al.* (2009) point out that HFACS categories are too wide and include error categories, which normally would be considered as performance shaping factors, e.g. system design. However, there is a human error by an engineer or a responsible manager behind many design-related or management-related PSFs. By using PSFs as a context, it is possible to trace an error to an individual.

Most taxonomies found in the literature are developed reactively by analysing the past incident statistics. This makes state of the art advancements more iterative but also more custom made for particular systems. Such process requires significant attention to uniformity of a sample as differences in system features should not be disregarded (Moors *et al.*, 2015). According to Dekker (2001), “*focus on superficial similarities blocks our ability to see deeper relationships, deeper patterns, deeper reasons and subtleties*” (p.253).

The retrospective approach depends on quality of past incident analysis reports and can vary across incident types (Madigan *et al.*, 2016). Focusing on past investigations also leads to exclusion of positive PSFs (Gibson *et al.*, 2009). It is important to note that human errors must be distinguishable from PSFs even when blame allocation is not an ultimate goal. Using an example by Gibson *et al.* (2009), where signage on gantry is not in line with signal sighting standards and is considered a design error. However, under older standards this signage might have been acceptable or there was a standard derogation issued for it thus making it a performance shaping factor.

The available human error taxonomies, despite considering causal factors, are still criticised for allocating a blame instead of identification of weaknesses in a system. Dekker (2001) claims that most taxonomies still follow “blame the human” approach without considering latent factors due to the illusion that classification is the same as analysis. He claims that only a few have attempted proper systematic review of human performance (in aerospace industry) through analyses spanning multiple years. Such research accounts for all the complexities between many processes and interactions that can contain PSFs.

It is should be assumed that systems are not safe (Dekker, 2001) and human operators are safety creating elements in the systems (Montano, 2011; Zeilstra and Van Der Weide, 2013). In this case classifications of past incidents by human error types are meaningless without understanding what led to these errors. It is investigation of the connections between “*tools, tasks, and operating environment*” (p.248) that is required (Dekker, 2001). Drivers’ everyday behaviour can be indicative of performance shaping factors embedded in a system (Salmon *et al.*, 2015). The pro-active identification of the PSFs embedded in a system can bring significant safety benefits but the reactivity of many taxonomies becomes a limiting factor for such studies. However, the existing taxonomies can provide a good starting point for this kind of pro-active assessment when those are used in context of unique design, driving task and organisational features.

With this thesis also envisaging an exploration of the system-embedded PSFs in day-to-day operations, the existing body of knowledge can contribute to identification of areas for analysis. The notion of design related factors (physical/technological environment) in HFACS is used for the research. However, the concept of performance shaping factors as a context for human errors is adapted from TRACEr-rail. This combination of concepts should allow examining metro drivers' performance from system point of view and identifying the relevant PSFs.

### *2.1.4 Performance Shaping Factors*

PSF taxonomies vary across the safety-critical industries similarly to the human error models. A range of PSFs taxonomies have been developed in the past decades for general applications (Hollnagel, 1998), nuclear industry (Mackieh and Cilingir, 1998; Kim and Jung, 2003), road transport (Stanton and Salmon, 2009), power plants (Boring *et al.*, 2007) and power distribution systems (Fereidunian *et al.*, 2010). Increased interest in human factors and quantification of human performance led to emerging of several PSFs taxonomies for railways.

Gibson *et al.* (2013a), in their work on Railway Action Reliability Assessment, produced a taxonomy of 27 PSFs named error producing conditions. Kyriakidis *et al.* (2012b) created rail-specific taxonomy of 43 PSFs which are divided in seven categories (Table 3). The taxonomy, called Railway PSFs (R-PSFs), distinguishes between dynamic and static factors that do or do not change with time. In comparison with other taxonomies, the R-PSF taxonomy takes into account interdependence between PSFs. It is also applicable to signallers, controllers and other operators within the railway network. To be versatile and reliable simultaneously, the taxonomy weights PSFs against each operator's duties (Kyriakidis *et al.*, 2012c). RSSB (2008) created a guide on human factors critical to human performance, which can be seen in Figure 4. Another human factors review by Gibson (2016) looks specifically into category A SPaDs.

| PSF category | Examples |
|---|---|
| **Organisational factors** | Training methods, safety culture |
| **Team factors** | Communication, coordination |
| **Personal factors** | Experience, motivation |
| **Dynamic personal factors** | Stress, fatigue |
| **Ambient factors** | Weather conditions, visibility |
| **System factors** | System design, HMI, working environment |
| **Task factors** | Workload, complexity |

**Table 3. R-PSF taxonomy. Adapted from Kyriakidis *et al.* (2012b)**

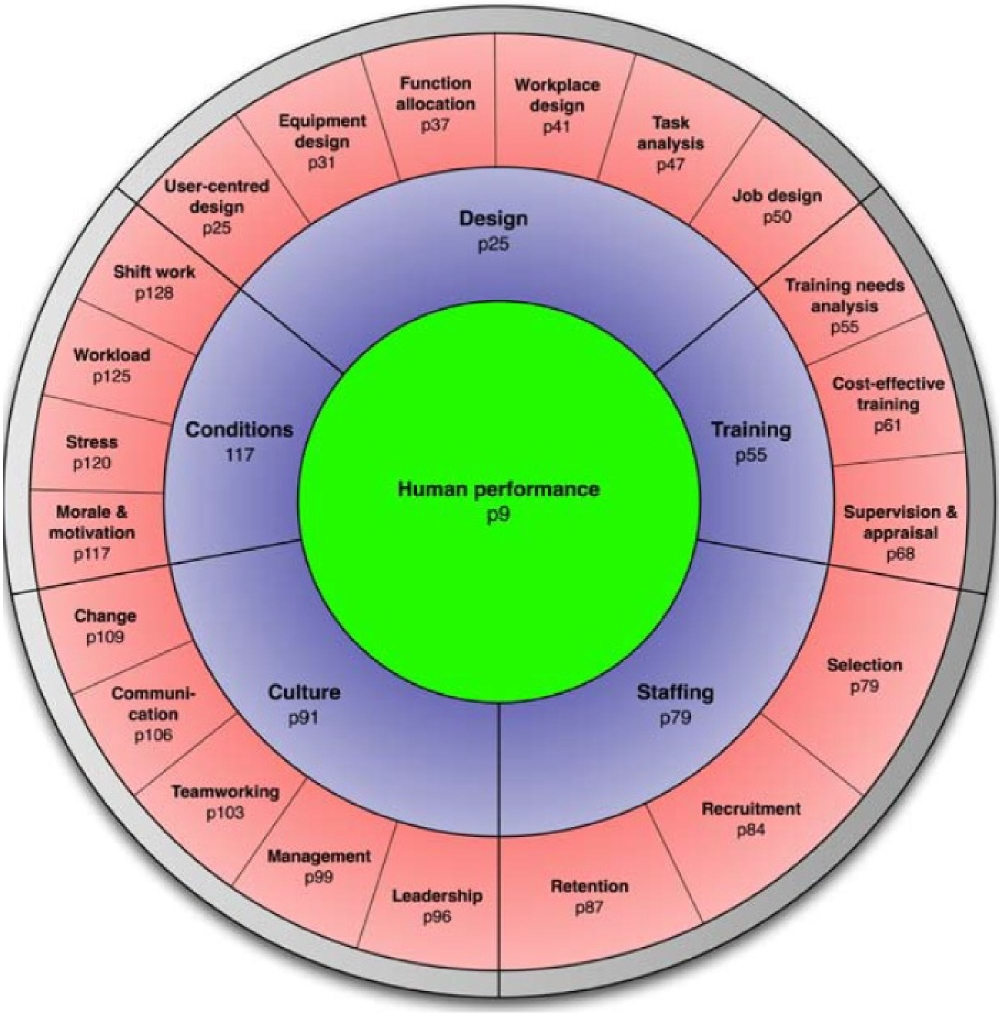


**Figure 4 Human Factors areas critical to human performance. Source: RSSB (2008)**

All of the taxonomies include design related factors but mainly focus on design deviations from accepted standards and misuse. Gibson *et al.* (2013b) claim that equipment "*can be a factor if it is not being used as intended, if it is faulty, if its* *design is not compatible with its use or if the layout is not in the order in which it is used*" (p.7). R-PSF research shows "system design" as top 3 contributing PSF in incident propagation (Kyriakidis, 2013). However, with SPADs still being high on researchers' agenda, there are concerns that this category solely focuses on signal sighting. Even though the taxonomies include design related factors, those are often vague (Dekker, 2001). Some design related factors can be logged as personal, e.g. misinterpretation of the DMI messages. Such error types might not account for a complex relationship between message design, concurrent tasks, workload and time pressure in a cab with poor desk layout. It is possible to observe some limited inter-dependence consideration in the taxonomies, e.g. Kyriakidis (2013) in his model does not assume relationship between system and environmental factors. This results in many PSFs potentially being left out and shows yet another drawback of retrospective approach.

The R-PSF framework is the most relevant for this research as it puts driver performance related incidents into an operational context. This is achieved through a notion of the inter-dependence of the PSFs. This is further explored in the follow-up work on the R-PSF framework (Kyriakidis *et al.*, 2015), where the authors began to investigate (albeit briefly) multi-factor relationships in incident propagation. Although the framework followed a standard human factors taxonomy development process (a retrospective analysis of past incidents, vague categories), it is significantly more advanced in holistically reviewing factors that contribute to train drivers' errors than the research presented above.

## 2.2 System design as a performance shaping factor

Nature and feature of the tools human operators work with can be both detrimental and beneficial to the system safety (Dekker, 2001). These tools create physical environment around the drivers in metro systems. Those can be located inside a cab, e.g. door controls, or be a part of a wider system design, e.g. platform mirrors. Research into influence of system design on train drivers' performance is often driven by high profile incidents or an attempt to understand and mitigate an elevated incident rates for a certain incident type. Minor incidents are often disregarded which does not allow addressing incident precursors (Madigan *et al.*, 2016). SPAD propagation is a research topic which has received a lot of attention in the past decades due to both reasons. For example, multiple investigations into Ladbroke Grove train crash tried to assess the driver's error from systematic perspective. In terms of system design, signal visibility (Cullen, 2000), complexity of track and signalling layouts combined with challenging environmental conditions (Lawton and Ward, 2005) were mentioned as contributing factors. Number of casualties and subsequent establishment of RSSB meant that the focus on SPAD incidents would continue. However, as mentioned in Chapter 1, the body of knowledge for other driver-related incidents is growing. According to Madigan *et al.* (2016), the contributory factor categories are consistent across the incidents related to train driving. However, it is unknown whether this consistency occurs when train dispatch tasks are also considered. The following sub-sections describe into research on rail incident types.

### *2.2.1 Signal Pass at Danger*

A report by RSSB (2004b) concluded that more research should be carried out to explore the role of infrastructure in SPaD incidents. According to Railway Industry Advisory Committee (2007) poor signal sighting design and layout of a cab are among the contributory factors in SPaD propagation. In order to address rising number of SPaDs, Naweed and Rainbird (2014) studied causal factors in this type of incidents across Australia and New Zealand. Signal sighting had been emphasised as the main issue, which is further induced by the density of infrastructure, e.g. overbridges can obscure signals.

Analysis of influencing factors in London Underground SPaDs showed distractors, attentional capacity, mental sets, signal sighting and design, sunlight and rail adhesion to have the biggest effect (Dray et al, 1999). Multer *et al.* (2015) stated that difference in rolling stock operated means problems with signal sighting from certain cabs, especially dwarf signals. They also claim that electric locomotives are at higher risk of SPaDs, partially due to more responsive controls and displays (assuming those are newer and have more DMIs). These advantages in train handling and in-cab design lead to train drivers spending more time on gazes inside a cab. For example, a more responsive controls (almost immediate acceleration) in electric trains means that the drivers should observe speed more carefully in situations with tight speed limits. On the other hand, Moors *et al.* (2015) demonstrated that currently, in terms of SPaD risks, train cab design is managed significantly better than workload or design of infrastructure, at least in New Zealand. However, they admit that there is still room for improvement in terms of the cab design.

Gibson (2016) wrote a comprehensive review on HFs in SPaDs across the UK railway network, showing knowledge, personal and management related incident factors (include human errors and PSFs) as the main causal conditions. Effects of work environment incident factors were significantly higher for passenger TOCs, whereas equipment incident factors – for freight. However, when a numerical risk factor in FWI (fatalities and weighted injuries) is assigned to each incident factor group, equipment is accountable for 26% of total risk. This investigation disregards differences between routes and systems. Moreover, Gibson's work, as many other SPaD focused research (e.g. Naweed and Rainbird, 2014; Multer *et al.*, 2015), puts emphasis on design of signalling when it comes to system design PSFs. Even though cab environment is often mentioned, no consideration is given how wider design of the physical environment affects signal sighting and interpreting. With PSFs being inter-dependent, high workload causing the SPaD could be created by a design feature, which has no association with signals.

### 2.2.2 Station overruns and failures to call

Station overrun are described as : "*Event in which a train which the driver is attempting to bring to a stand at a booked station stop proceeds beyond the designated stopping point such that any door intended to be available for public use at that station is no longer on the level platform" (Association of Train Operating Companies, 2006).* Failure to call is "*Failure of a train to make a booked station stop in cases where the driver has made no attempt to apply the brake*" (ATOC, 2006). From the definitions it is possible to see that these two incident types are quite similar with only difference being whether the brakes were applied or not. ATOC (2006) guidance also suggests that in an event of a station overrun, when a red signal was passed, to investigate this incident as SPAD. RSSB (2010) found strong correlation between drivers with a station overrun history and SPaD occurences. Evans (2005), in his paper on railway risks, includes overruns in the same category as derailments and collisions because of their particular importance and high profile.

Station overruns are highly correlated with low rail adhesion (LRA) conditions which hinder braking performance. There is significant literature explaining the phenomena of the LRA (e.g. Office of the Chief Investigator, 2009; Ward *et al.*, 2012; Zhu *et al.*, 2012) but in the UK it is mostly a combination of fallen leaves and morning/evening dew and fog. Light rain can also contribute to these conditions. In terms of system design, track design and geometry were found to have little influence on incident propagation (Office of the Chief Investigator, 2009). However, in case of a new rolling stock or newly qualified drivers, unrealistic expectations of braking performance based on the “dry running” experience were found to be a causal factor. Another similarity between SPaDs and station overruns is that signal disregard can be a human error causing both incidents (Railway Group Standards, 2000). Hence many PSFs applicable to SPaDs potentially can be transferred to station overruns and failure to call incidents.

RSSB (2010) in their analysis of station overrun and failure to call statistics, concluded that memory errors are the main mechanism triggering such incidents. The main PSFs are personal factors, e.g. physical health, fatigue etc. The environment factors (temperature, lighting) only involved in less than 5% of incidents where PSFs were identified. This, however, excludes the LRA. Data quality is unknown in this retrospective assessment as it covers the period between 1999 and 2006 where simply blaming drivers during incident investigation was a norm.

Research on the international level mostly focused on technical solutions to mitigate these incidents. Sato *et al.* (2010) proposed a system that prevents station overruns by supplying a driver with information on a predicted stopping position. The position is calculated using vehicle velocity and deceleration. Information can be supplied using the DMI (Inoue and Yamamoto, 2006). Train Automatic Stop Control (TASC) systems usually are coupled with Automatic Train Operation (ATO). Even when automation is not considered, the cost of retro-fitting of such driver advisory systems is high. Moreover, taking some of the core options of train drivers away from them can cause situations of mental underload and low arousal levels. This can be even more harmful to human performance, as shown by some wrong side door activations (Connor, 2007). Providing a concurrent advice to drivers on station arrivals is associated with other issues. The in-cab DMIs are known to distract from front monitoring (Davies *et al.*, 2007), which, in this case, will be combined with a driving task already causing increase in mental workload (Myrtek *et al.*, 1994).

### 2.2.3 Wrong side doors activations

Metro systems are usually equipped with passenger-controlled doors on both sides of a car, i.e. when a passenger has to press a button to open a set of doors. However, the doors on one of the sides have to be first released by a driver. Wrong side door activation incidents occur when a driver releases an incorrect side of doors. Even though there is an additional defence layer with the passenger activated doors, the risk increases with passenger loadings. In a crowded train there could be some passengers leaning on the doors and activating the open buttons involuntary. This creates a risk of people falling on a track and being trapped or hit by a train on the adjacent tracks. On mainline railways some platforms are shorter than trains. This requires a selective door activation procedure complicating the task even further.

Halliday (1995) investigated the issue in a hope to address these incidents on Thames Train Class 165/166 rolling stock. It was found that the problem is wide-spread among the companies using the DOO trains. Halliday (1995) identified a left-hand bias in the railway industry, when most of the platforms are to the left from the drivers. Hence a door opening procedure becomes an action on the skill-based level and a non-routine encounter of a right-hand side platform might not activate a different action. It is also noted that the door opening task competes with other station stop tasks and design interventions splitting the tasks in time would have the biggest positive impact. Basacik and Gibson (2015) claim that "*using unofficial cues on the non-platform side rather than official stopping markers*" (p.444) can lead to the wrong side door activations through decision errors. Among the PSFs, equipment design and layout were highlighted the most in their analysis, e.g. platform side change, a location of in-cab CCTV, poor visibility of stop markers on island platforms, platform signals being on opposite side of the track. However, drivers' door opening performance is much better than the one predicted by human reliability assessment techniques thus serious improvement more likely to be achieved with technical solutions (Basacik and Gibson, 2015). The technical solution offered was Correct Side Door Enable (CSDE) technology, used by London Underground. It is a platform-based system of beams, which blocks activation of an incorrect side of doors.

Connor (2007), looking back at his experience as a London Underground driver, states that the repetitiveness of a train driver job causes a mind drift and lose of awareness about the current station. Even though technical solutions exist, e.g. the CSDE, drivers still manage to override those and create an incident (Connor, 2007). ND BOMEL Ltd (2009), in assessment of a risk of an automatic selective door operation system, mostly focus on inadequate system designs and overreliance of such systems on a correct stopping position. RSSB (2011a) looked into introducing the selective door operation and emphasised the need for consistency across the system if this approach is to be introduced safely. This could be also applied to consistency in terms of platform sides.

Overspeeding

Although current UK risk estimation models assign relatively low fatality and injury risk to such incidents (Monk *et al.*, 2015), overspeeding often is a major cause of train derailments around the world. It can lead to very high profile accidents and rail crashes. Several high profile overspeeding accidents are summarised in Table 4.

| Where | Date | Allowed/Actual speed | Casualties | Injured |
|---|---|---|---|---|
| **Spain** | 24/07/2013 | 80/190 km/h | 79 | 130 |
| **China** | 28/04/2008 | 80/131 km/h | 72 | 416 |
| **Japan** | 25/04/2005 | 70/116 km/h | 107 | 562 |
| **Australia** | 31/01/2003 | 60/117 km/h | 7 | No information |

**Table 4. Examples of train crashes due to overspeeding in XXI century.**

Baysari *et al.* (2008) studied two samples of accident investigation reports from the UK and Australia. They have found that speeding problem is more acute in Australia although not non-existent in the UK. They also discovered that majority of the speeding incidents are violations of driving rules, i.e. deliberate human actions. Furthermore, 70% of violations are routine to drivers who got used to such style of driving. Often drivers will be speeding to make up for a time lost early on route. According to the UK mainline train drivers, surveyed by Monk *et al.* (2015), main PSFs are lack of knowledge, fatigue, memory lapses, attention lapses and an incorrect design and placing of the speed restriction signage. However, other PSFs involved are distractions, layouts and visibility, lighting conditions, late running etc. The results also suggest that there is a scope for redesigning speed restriction signs to improve their visibility (Monk *et al.*, 2015). Blanchard and Lowel (2009) in their incident investigation case study showed that clustering of trackside infrastructure causes a higher workload and task demands. Moreover, co-location of signs and signals can create conflicts due to divided attention demands and conflicting speed information (a signal at caution advises a lower speed, whereas a speed limit allows high speed running). According to Clements *et al.* (2011) distractions such as the eco-driving HMIs can lead to overspeeding, but the risk is rather low.

### 2.2.4 Passenger entrapments

The platform-train interface is the most hazardous area in a modern railway system in the UK (RSSB, 2016). Majority of the fatal boarding or alighting accidents on UK railways happen due to a victim falling in a gap between a train and a platform

(RSSB, 2011a; RAIB, 2012). Hence a significant research focus on mitigating the risks associated with platform gaps, e.g. Rajkumar *et al.* (2012). In many cases a driver or a dispatcher has no means of preventing an accident (RAIB, 2013). However, if a person is entrapped by a door closing sequence, whoever is responsible for the safe train dispatch should notice the incident and prevent it. In DOO systems it is usually a train driver. Although the modern train doors are equipped with anti-entrapment sensors (cycling the doors when an entrapment is identified), the drivers sometimes rely too much on the equipment thus only briefly check the dispatch equipment (RAIB, 2016a). Even though RAIB (2016b) provides a proof for the decreased PTI risk in the UK metro systems, approximately a half of the RAIB investigated trap & drag accidents occurred in London Underground or Tyne & Wear Metro. Even higher proportion of such investigations involves the DOO stations thus strongly suggesting elevated risk levels. However, Cynk *et al.* (2015) advocate that "*DOO led dispatch improves performance at the PTI*" (p.47) but mention that an evidence base for this claim should be improved. On the other hand, in their older assessment of the DOO dispatch, Basacik *et al.* (2009) present the DOO systems as prone to SPaDs and additional entrapment risks.

According to RSSB *et al.* (2006), a composition of passengers (by age and mobility), their behaviour and crowdedness influence risks at the PTI. Moreover, they concluded that design-related factors are mitigated well by the UK railway standards. Cynk *et al.* (2015), in their review of 171 PTI incidents, also agree that the passenger characteristics are an important causal factors. Their research suggests visibility and obstruction issues as contributory factors too but it does not distinguish between DOO and train/platform staff dispatch. Cynk *et al.* (2015) claim that, in terms of the DOO dispatch, platform cameras support better PTI performance than mirrors due to consistency in picture size. Swanson (2004), in his view on advanced light rail communication system, writes that external door cameras are much more effective than rear or platform mirrors, especially in a multi-car operation and at curved platforms. However, poor design of the CCTV cameras can be a serious risk factor not allowing a DOO driver to identify the entrapments (Traub and Fraser, 2013; RAIB, 2016b).

Technical solutions offered to tackle the problem are based on the use of the CCTV equipment. RSSB (2014a) explored using image recognition software in order to automatically detect hazardous events at the PTI. Traub and Fraser (2013) demonstrated that a zoom option can bring benefits to safe dispatch at CCTV DOO locations. CCD Design & Ergonomics Ltd (2005) explored using more in-cab CCTV images to extend DOO onto longer trains. The research also indicates a necessity to correlate allowed dwell times with passenger levels. No discrepancies were found between performance in daytime and night time. RSSB (2007a) investigated human factors in context of CCTV monitoring and DOO driving.

There is not a lot of research done on the DOO platform mirrors even though those provide rather cost effective solutions well fit for certain systems. RSSB (2014b), in their outline strategy for converting the UK rail network to the DOO, do not consider platform mirrors. Instead, the CCTV and on-board rear-view mirrors (for low traffic routes) are discussed. Research on the CCTV DMI design does not take into account potential negative effects on other tasks, e.g. the correct side door release (Basacik and Gibson, 2015). Research on other design-related PSFs is fragmented too. According to Cynk *et al.* (2015), certain platform designs with limited covered area cause overcrowding at certain parts of a platform thus increasing the risks.

### *2.2.5 Metro-specific research*

The above review of the available literature for each incident type clearly indicates scarce nature of studies focusing on urban rail. Although complexity of train driving in urban environment is acknowledge in the literature (Naweed and Balakrishnan, 2014), only rare studies address it. Among the incident types explored in such studies are SPaDs (Dray *et al.*, 1999) and wrong side door activations (Basacik and Gibson, 2015). The passenger entrapments are often approached from passenger composition and behaviour points of view (RSSB *et al.*, 2006; Cynk *et al.*, 2015), which can be partially considered relevant to metro systems. It is possible to notice that SPaDs are not predominant research area in this sample of studies but this could be caused by small sample size.

## 2.3 Physical environment factors

Drivers' immediate physical environment, i.e. cab environment, is a significant contributor to the drivers' performance (RSSB, 2012).  It is associated with numerous environmental factors which can be further induced by inappropriate design of a cab. Moreover, the cab has to be designed with the end user in mind in order to be ergonomically sound. Van Der Weide *et al.* (2013) described successful application of a user-centred design in procurement of the new Amsterdam Metro rolling stock. Early and continuous involvement of the metro drivers led to a better quality and acceptance in this project. The cost of redesigning or making amendments to the existing designs is hundred times more than cost of changes during the design phase (RSSB, 2008). In spite of that, a growing understanding of the physical environment influence and many rolling stock retro-fitting projects have sparked numerous projects looking on design of different parts of railway systems in the past decade.

### *2.3.1 Noise*

Design of the cab environment was studied by Johnson *et al.* (2009) in order to reduce heavy rail train drivers exposure to noise. They concluded that drivers are exposed to a lot of low-frequency noise what is not in line with the current Federal Railroad Administration (US) rules. Authors proposed several active and passive measures e.g. better door seals, removing air gaps, vibration damping sheets, and unwanted sound cancellation using microphones and speakers. Decreasing drivers' exposure to noise is important as it is a causal factor for annoyance, drowsiness, stress, tiredness and fatigue (Sümer *et al.*, 2006; Johnson *et al.*, 2009; Maguire, 2009). Besides, the low-frequency noise is known to cause cases of chronic insomnia and depression (Mirowska and Mroz, 2000) as well as errors in judgement (Maguire, 2009). Background noise has been found an important environmental factor in SPaD propagation (RIAC, 2007). European regulations for noise levels do not account for harmfulness of the low-frequency noise, which is dominant in a driver cab. The proposed using of A-weighting scale means that the low-frequency noise conditions are treated as the least harmful (Maguire, 2009).

### *2.3.2 Vibrations*

Exposure to vibrations for more than 2 hours increases incident risks due to negative effect on train drivers' arousal levels (Keun Sang and Ohkubo, 1994; Yang *et al.*, 2012). However, sometimes engineering solutions to issues can be rejected. Ahmadian and Vahdati (2006) provided an example of improved vibration levels that can cause more discomfort to a driver than the original design due to noise and body frequency coinciding.

### *2.3.3 Monotony*

Yang et al. (2012) found that train drivers' working environment is more simplistic thus more monotonous. Thiffault and Bergeron (2003) reviewed some proposals to the highway design in order to fight monotony of the environment, including design inducing mild stress in drivers, addressing a problem of "too good roads", curves and kilometre pegs to induce visual stimulation, and even road-side art. As the mainline railways become more standardised, moving towards closeness of a system observed in urban rail, monotony of the task increases (RIAC, 2007). This, along with improved comforts offered by cab environment, contributes to reduced concentration levels.

### *2.3.4 Thermal conditions*

Temperature is one of the most important environmental factors as human operators have only a narrow band of tolerance for changes in temperature before it starts affecting how they carry out a task (Elliott, 2012). Thermal conditions have been found an important environmental factor in SPaD propagation (RIAC, 2007). Heat can have adverse effects on human operators causing unbearable and sometimes dangerous conditions in the cab (Ružić and Časnji, 2011). In their research on tractor driver cabs, Ružić and Časnji (2011) have showed that cab design can have competing goals of increasing visibility (increasing glass) and reducing adverse effects of solar radiation. The later usually being ignored in favour of improved ergonomics and visibility via bigger windshield. Authors have proposed several mitigation measures but some of those are not applicable to rail operations. The newer rolling stock is usually equipped with air conditioning units.

### *2.3.5 Lighting*

Such conditions as fog, rain, snow and direct sunlight are known to cause brightness and contrast distortion, and change drivers' perception of a distance (RIAC, 2007). This can significantly affect observational tasks and actions taken by train drivers. Glare is very hard to manage and design for but it affects both the lineside and in-cab signalling. In case of the lineside signalling the positive fact is fixed location of the signal hence it is possible to introduce anti-glare measures to improve signal sighting. The in-cab signalling, on the contrary, is always in motion and it is almost impossible to model when a screen will be affected by the sunlight. Thus it is necessary to increase brightness of the screen to compensate for the sunlight. However, too bright screen becomes a problem during night-time or tunnel operations when it becomes main source of illumination in a cab. During the trials on Cambrian line the DMI screens caused a discomfort among drivers because of extreme brightness at night (RSSB, 2012). On the other hand, it is possible to use lighting conditions as a positive PSF. Ceci et al. (2013), in their appraisal of a tunnel design, showed that introduction of lighting features in the infrastructure design allow increasing arousal levels and visual stimulation.

## 2.4 Other factors

Madigan *et al.* (2016) showed that different incident types can be propagated by similar factor categories. Even though these categories can be broad, there is a scope for research of the PSFs which are common across different incidents. Previous research on the PSFs not related to design shows similar factors causing station overruns and failures to call (Roels and Mills, 2010), SPaD (Gibson, 2016), wrong side door activations (Basacik and Gibson, 2015), and overspeeding (Baysari *et al.*, 2008; Monk *et al.*, 2015). Among these factors are boredom, workload and stress, memory and attention lapses, distractions, various organisational factors. All of these PSFs are not unique to railways and are known latent conditions in other safety critical industries.

Some of the other factors mentioned above can be produced by a single design feature or a combination of those. Hence addressing a problematic layout, e.g. a location of the DOO equipment, might bring significant safety benefits to tasks not related to train dispatch. However, it is necessary to identify which incidents have similar causal factors in the urban rail systems first.

## 2.5 Conclusions

The body of knowledge related to the effects of system design on train drivers' performance is increasing. However, it is still rather fragmented and disjointed even though common PSFs have been identified in various tasks. With researchers selecting a certain human error as an avenue for an investigation, complexities of the full driving task tend to be not accounted for. Research often favours a categorisation approach, meaning that there are still risks for the focus being skewed towards human errors rather than latent conditions. Moreover, the available taxonomies can oversimplify causal factors to achieve better fit to more systems. Hence deeper incident underlying structures are overlooked and inter-dependence of the PSFs is often ignored or addressed only partially. Understanding an interface between different motivations, pressures, latent conditions and priorities in train driving can be a significant step forward towards a systematic approach to train driver errors.

The discussed research is usually retrospective and is significantly skewed towards high profile accidents. This is understandable as those accidents usually have the best quality of data when it comes to investigation reports. Hence there are significantly more studies into human performance, human factors and PSFs in SPaDs than in the passenger entrapments or the wrong side door activations. Both the passenger entrapments and wrong side door activations are highly related to urban rail systems and DOO dispatch methods. The incidents associated with station procedures have been explored less from human factors point of view despite posing significantly higher risk to passengers. This creates a gap between research focus and the actual passenger risks.

An aspiration for bigger and better data samples in a study can lead to grouping of incidents from various rail systems. Such grouping does not allow exploring the findings in context of a specific train driving process unique to each system. However, the incident propagation across the systems can be caused by different PSFs. The concerns are about how well these differences are captured in incident investigation reports and subsequent studies. Metro systems are rarely featured in this research highlighting a certain disregard for the differences with mainline railways. Only a small number of urban rail specific research has been identified but some PSFs are identified in RAIB case studies. Moreover, metro-specific research has not been found for certain incident types, e.g. overspeeding.

The reactive approach of many studies can also be a limiting factor as the basis for research is created by people (incident investigators) with sometimes a limited exposure to the state of the art in rail PSFs. Moreover, the data base itself is limited when some incident types are considered. Hence opportunities for uncovering latent conditions not known before are poor. On the other hand, ever-increasing technical complexity of the rail systems creates more design related performance shaping factors which are present in everyday operations. However, these factors require more complicated structures and alignments to cause incidents due to better system defences. This is why more pro-active approach to the train drivers' safety performance issue should explore everyday behaviour of drivers' population in order to uncover unknown PSFs, particularly for urban rail systems where the literature is scarce.

### *2.5.1 Literature review gaps and research questions*

The presented literature review identified some gaps in the body of knowledge related to urban rail systems which contributed to establishing the research questions in Section 1.5. The paragraphs below provide a brief overview of how the literature review links to each of the research questions.

1. *Is the existing mainline research applicable to urban rail systems?*

The literature on human performance presented in this chapter clearly defines a range of factors or factor categories affecting train drivers' performance. Although urban rail systems can have many elements common to mainline railways, the notion of inter-dependence of PSFs implies that a change in one factor might affect how other factors interface with drivers. Only a few of the research works presented above are focused on urban rail systems thus raising the question whether the previous research in rail human factors can be extrapolated to urban rail. It is important to understand how (and if) individual characteristics of a system change the way how PSFs affect drivers. Chapter 4 shows that there are many characteristics of urban rail systems that are different to mainline railways and thus need to be taken into consideration before applying the previous research. For example, Driver Only Operation is more wide spread in urban rail and station stops are far less frequent in mainline railways.

2. *What design features act as performance shaping factors in metro driver related incident propagation?*

The existing body of knowledge identifies some design features as PSFs in mainline rail, e.g. track and signalling layouts (Lawton and Ward, 2005; Multer *et al.*, 2015) or a cab design (Davies *et al.*, 2007). However, a potentially different risk profile in urban rail systems (as discussed in Section 1.1) justifies an in-depth exploration of other parts of the infrastructure, e.g. station environments, in comparison to the SPaD-driven research. Nevertheless, well researched design elements should be also explored but in a context of local operational features, e.g. a lower line speed.

3. *What design and procedure modifications to existing urban rail systems can mitigate the effects of these PSFs?*

It has been shown that certain design features are managed relatively better than other design features (Moors *et al.*, 2015). For example, the adverse factors increasing the risks of wrong side door release can be mitigated through certain technical interventions (Basacik and Gibson, 2015). Similarly, design and procedure modifications can be introduced to limit or remove metro drivers' exposure to some adverse PSFs. However, as with the findings on the effects of PSFs, the modifications cannot be simply transferred from mainline railways. For example, improvements to a communication process between a driver and a guard cannot be applied in urban rail systems as this factor rarely exists in such systems.

# Chapter 3. Research methodology

## 3.1 Introduction

The prior understanding of the isolated system design factors discovered in railway industry up to date has been formed through the literature review in the previous chapter. Chapter 2 allowed identifying the avenues for further research and formulating research design suitable for addressing the research questions. This chapter presents the research methodology in context of philosophical paradigms and research tools along with advantages and limitations of those.

Although previous research establishes that railways are complex socio-economic systems (Naweed and Balakrishnan, 2014), it rarely explores the underlying factor relationships in-depth. Past research has often resorted to categorisation and generalisation in order to improve transferability of findings across the systems and countries. However, such approach also leads to focus on the most evident factors and human failure mechanisms only.

This thesis acknowledges the concept of inherit unsafety of systems as presented by Dekker (2001). In his "*Re-invention of Human Error*", he advocates that human operators are tasked with creating safety in systems where design, task and operational environment factors inter-connect in ways that can lead to performance failures. Dekker (2001) claims that it is important to study how people successfully operate the systems along with the factors causing the failures. The importance of understanding everyday human actions, albeit in different situations, had received support in the literature (Montano, 2011; Zeilstra and Van Der Weide, 2013; Salmon *et al.*, 2015). In terms of urban rail, these actions usually focus on visual managing of physical environment and the train itself (Naweed and Balakrishnan, 2014).

## 3.2 Research philosophy

In order to truly address the complexity of metro systems, the research methodology should satisfy a number of parameters. As an engineering thesis, it should be objective in order to investigate causality and create findings transferable to other systems. The objective methodologies usually involve quantitative methods that are undertaken in unbiased environment (Philips and Burbules, 2000; Creswell, 2014). As the body of knowledge predominantly consists of studies employing quantitative methods, such tools are important precondition for comparison with the state of the art findings.

An aspiration of research in everyday metro drivers’ performance calls for more naturalistic research design though. The naturalistic research studies subjects in their natural setting. For metro drivers such setting is their normal work environment. The biases are unavoidable in naturalistic research (Eugene *et al.*, 2015) and those can be present in a form of various factors, e.g. operational conditions. It does not automatically invalidate the numerical research but the spatial and temporal context of drivers’ jobs, operational environments and physical environment needs to be understood in order to pinpoint those biases. Moreover, some groups of performance shaping factors, e.g. latent conditions associated with safety culture, will have a significant subjective side to those. This is why subjective research methods are equally important to this thesis. Such methods allow human decisions to be studied on a micro level (Mangan *et al.*, 2004) and context to be provided to findings through iterative use of various qualitative methods (Crotty, 1998).

Traditionally, there has been a clear division between quantitative and qualitative methodologies in terms of governing paradigms, known as positivism and constructivism respectively (Creswell, 2014). These research philosophies or organising structures revolve around researchers’ epistemology (theory of knowledge) thus often prescribing selection of methods (Mangan *et al.*, 2004; Feilzer, 2010; Creswell, 2014). This, however, goes against an aspiration for the multi-dimensional thesis which investigates the research questions through different data levels.

A method-driven pragmatic approach is selected for this thesis. Instead of being constrained to methods prescribed by research philosophies, such approach addresses a task at hand and selects tools fitting the most. Creswell (2014) indicates that pragmatism allows combining qualitative and quantitative techniques without committing to any philosophy. Such triangulation “*can overcome the potential bias and sterility of single methods approach*” (p.569) (Mangan *et al.*, 2004). Pragmatic view on the investigated world is of the “*world with different elements or layers, some objective, some subjective, and some mixture of the two*” (p.8) (Feilzer, 2010). Hence the pragmatism is the only paradigm being able to facilitate research where quantitative findings have to be enriched by subjective explanations and contexts.

## 3.3 Research design and methodology

With the research objectives outlined in Section 1.5, the thesis progresses towards more detailed and specific findings with each chapter. An iterative strategy of inquiry is selected to support this. The strategy includes combination of deductive and inductive elements where each step refines findings of the previous step but also generates new theories if necessary (Bryman, 2014). The sequential design (Creswell, 2014) allows dividing the research process into several phases as seen in Figure 5. Feilzer (2010) praises such sequential design for flexibility to adapt the consecutive methods to fit the findings of the previous phase.

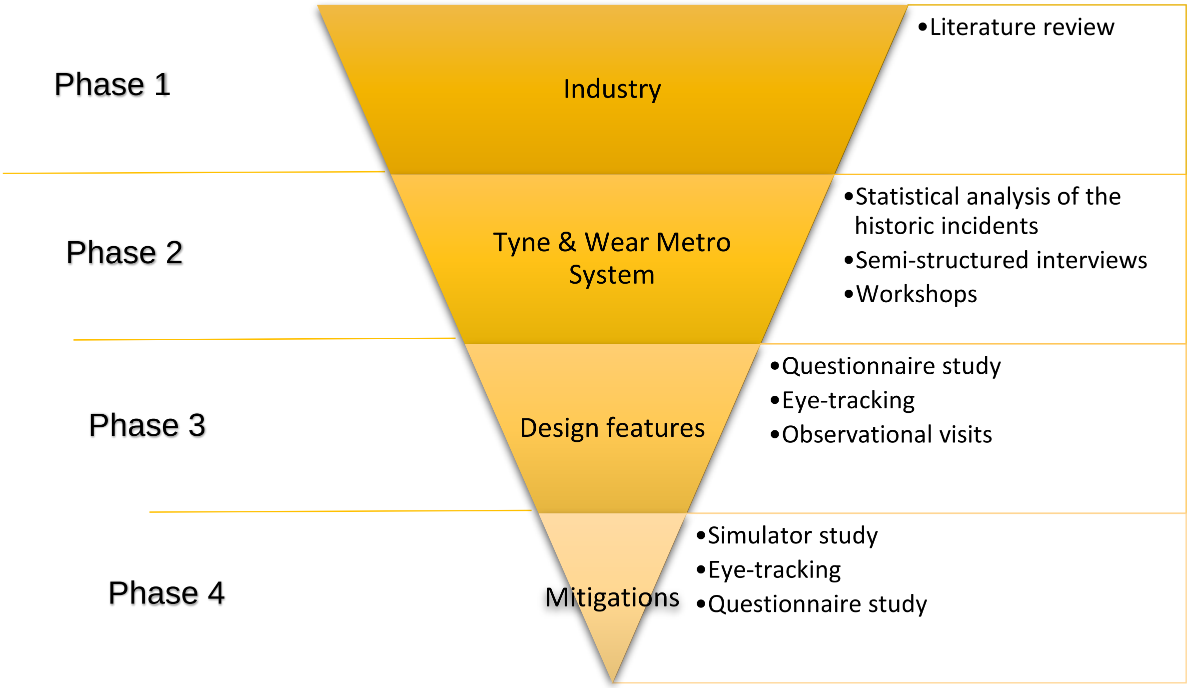


**Figure 5. Phases of the thesis with research levels and respective methods**

The research design proposed also falls into multilevel mixed data analysis category (Teddlie and Tashakkori, 2009) with one exception. It mixes the quantitative and qualitative techniques within numerous phases adding a concurrent element to the design. This allows the methods to supplement each other producing a more holistic understanding of each research level. The structure of the thesis is shown in Figure 6.

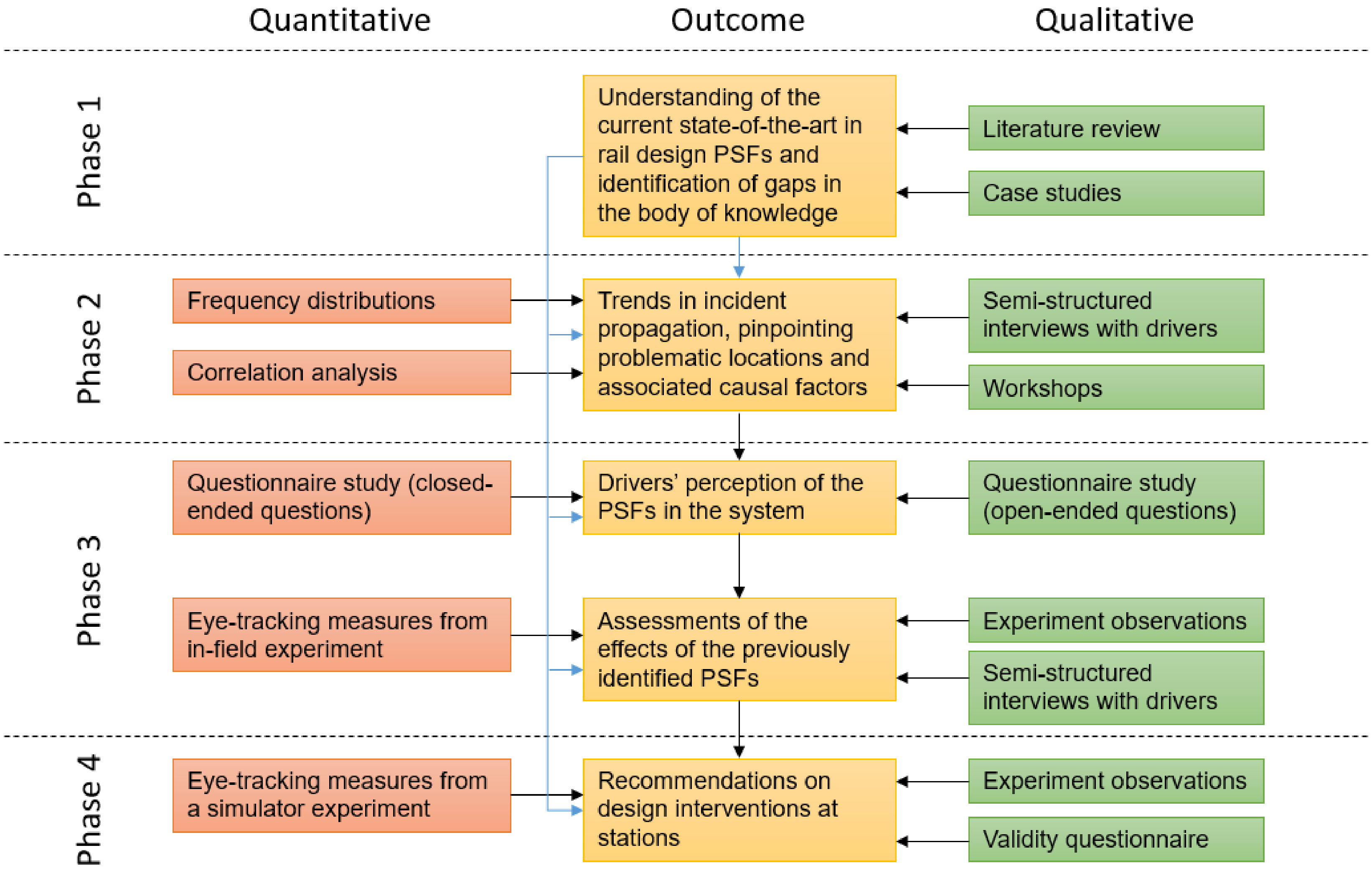


**Figure 6. Structure of the thesis showing the combination of sequential and concurrent designs**

The literature review is the basis for Phase 1 where the current state of the art related to PSFs in railway industry is identified. The predominant methodologies used at this level are qualitative methods, including case studies of incident investigations. The results of this phase are fed to all the next phases enriching the discussion on results and allowing answering the research question 1.

In Phase 2 the research starts investigating the case of Tyne & Wear Metro through the numerical analysis of past incidents statistics. It allows identifying design features potentially associated with the design PSFs. However, it is qualitative methods that facilitate detailed and in-depth study (Patton, 2002). In Phase 2 such tools as semi-structured interviews and workshops with Tyne & Wear Metro drivers complement the findings of the statistical analysis. Those also ensure that standardised measures of quantitative methods do not discard varying perspectives and operational context (Patton, 2002). This combination of methods enables progressing the research to a more detailed level where individual PSFs are considered.

Phase 3 is where the majority of the naturalistic research is conducted in this thesis. Firstly, the drivers are asked to assess their experiences in a number of real-life scenarios focusing on the previously selected design features. The developed questionnaire allows quantitative analysis along with open-ended questions which give participants an opportunity to explain their perspectives. Secondly, an in-field eye-tracking experiment is designed as a naturalistic driving study. The experiment measures relative cognitive impact of certain design elements, which has been highlighted before as an area requiring in-depth exploration (Naweed and Balakrishnan, 2014).

The eye-tracking experiment provides objective dimension to a more subjective questionnaire study. However, the interpretation of the eye-tracking findings is rather subjective and uses the experiment observations and the semi-structured interviews to generate final results. The flexibility of the sequential design becomes a serious advantage in Phase 3 because the methods are designed taking into account the previous phases.

Subjective assessment of the in-field eye-tracking results requires hypotheses to be checked with less naturalistic methods in Phase 4. Hence a simulator study is conducted in order to test the proposed mitigation measures in a controlled environment where tasks can be manipulated according to the research objectives

(Eugene *et al.*, 2015). The research in this phase focuses on specific elements of the station design, associated PSFs and proposed mitigations. The ultimate goal of this stage is to provide a set of measurable actions which can be implemented in other similar systems in the future. This phase mixes the methods too but not to the same extent as the previous stages. The qualitative techniques are mostly used for the simulator validation. However, the objective findings would be questionable without this validation.

# Chapter 4. Tyne & Wear Metro system

Tyne & Wear Metro is the application case used throughout the entire thesis as most of the experiments involve its drivers and their performance. The analysed system provides varied operational environments as it uses a combination of a purpose built, legacy and mainline infrastructures. The design features encountered in Tyne & Wear Metro are typical for both urban and sub-urban railways. Such mix makes this system perfect for assessment of the system design PSFs as it can be done through comparison within the network.

## 4.1 Background to the research

Motivation for this research emerges from issues, e.g. spikes in driver-related incidents, encountered by the operator of Tyne & Wear Metro (DBTW) at a time, which could not explain the root causes. Newcastle University was approached to assist in exploring the mechanisms triggering these types of incidents. The research proposed in this thesis is the result of this collaboration.

## 4.2 Overview of Tyne & Wear Metro

Tyne & Wear Metro is located in the Tyne & Wear conurbation that connects Newcastle upon Tyne, Gateshead, South Tyneside, North Tyneside and Sunderland. It first opened in 1980 and mostly adapted existing heavy rail infrastructure. Today the system spans more than 78 km and has 60 stations. A map of Tyne & Wear Metro can be seen in Figure 7. The system has two routes; the South Gosforth to Pelaw section of the network is considered the “core” of the system, as both routes pass through it, and thus it has the highest daily throughput of trains. The system carries more than 40 million passengers per year (Department for Transport, 2016) It is publically owned but operations are franchised out for the period between 2010-2017, which includes this thesis’ time span.

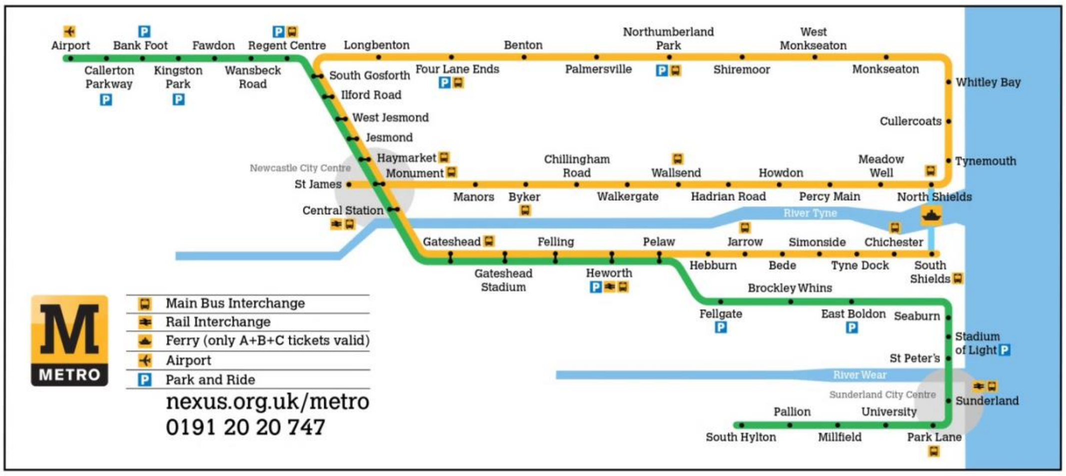


**Figure 7. Map of Tyne & Wear Metro.**

## 4.3 Fleet

The fleet consists of 45 two-car train sets. Tyne & Wear Metro uses the original class 994 rolling stock, which underwent its ¾ life refurbishment between 2011-2013. The cab layout is rather unusual with the cab taking only half of the width of a metrocar. The other half is used for passenger seats. The refurbished trains have similar cab design with a modernised driver seat and a new driver advisory system (DAS) called FASSI. A Heating, Ventilation and Air Conditioning (HVAC) unit had been installed during the refurbishment but was taken out on the later date due to drivers' complaints. More minor changes, which have a potential to influence drivers' performance, are described later in the thesis.

## 4.4 Stations

The majority of the stations in Tyne & Wear Metro system are located overground. There are only 8 underground stations in the network (St James, Monument, Manors, Jesmond, Haymarket, Central, Gateshead, Sunderland). However, Tyne & Wear Metro's own classification counts built-over (subsurface) stations, such as Regent Centre, as underground stations. Using Tyne & Wear Metro's own classification method, 13 stations in the system can be considered underground (the previous 8 plus North Shields, Four Lane Ends, Park Lane, Regent Centre, Heworth). Many overground station are "legacy" station adapted from the older heavy rail system.

There are 7 line and service terminus stations in Tyne & Wear Metro. Line terminuses (Airport, St James, South Shields and South Hylton) have either a single platform, or a layout allowing trains to arrive at any of the two available platforms. The service terminuses are used for short services and have turn-back facilities at a station, or in sidings. Many more stations have turnback facilities, which are used in events of disruptions or during line closures. The design of the stations, especially the ones purpose-built for Tyne & Wear Metro, is rather standardised.

The majority of the stations fall into one of the three types of standard designs used in Tyne & Wear Metro (Figure 8). Type 1 stations include most overground stations, all built-over stations and two underground stations. The underground Type 1 station differ by use of monitors instead of mirrors as dispatch equipment. 2 of the line terminus stations use Type 1 design but only with 1 platform. Type 2 stations are only overground stations with an island platform. These usually include some ticketing/waiting canopies on a platform and mirrors as dispatch equipment.

However, Airport as a Type 2 terminus station has monitors. Type 3 designs can only be encountered at the Newcastle and Sunderland city centre underground stations. Table 5 summarises what category is applicable to each station.

All of the design types include Driver Only Operation dispatch equipment (a mirror or a monitor), platforms, and a running signal. There are stopping position indicators on platforms, which usually span at least 1 metre to provide some flexibility to the drivers. Instead of the stopping position markers, Type 3 stations include a small sign on a left hand side wall. Despite a high level of standardisation, some of the stations deviate from these designs. For example, Tynemouth (Type 1) and Northumberland Park (Type 2) have a running signal on the opposite side of the track to where it usually is in respective design types. On the contrary, Cullercoats (Type 1) has no running signal at all.

Apart from the predominant DOO dispatch method for each design type, the stations also differ in the platform width. Type 3 stations, due to space constraints and being underground, feel more confined whereas the “legacy” stations tend to have very wide platforms. Not only the general feel of the stations is affected by this but also passenger visibility on the approaches.

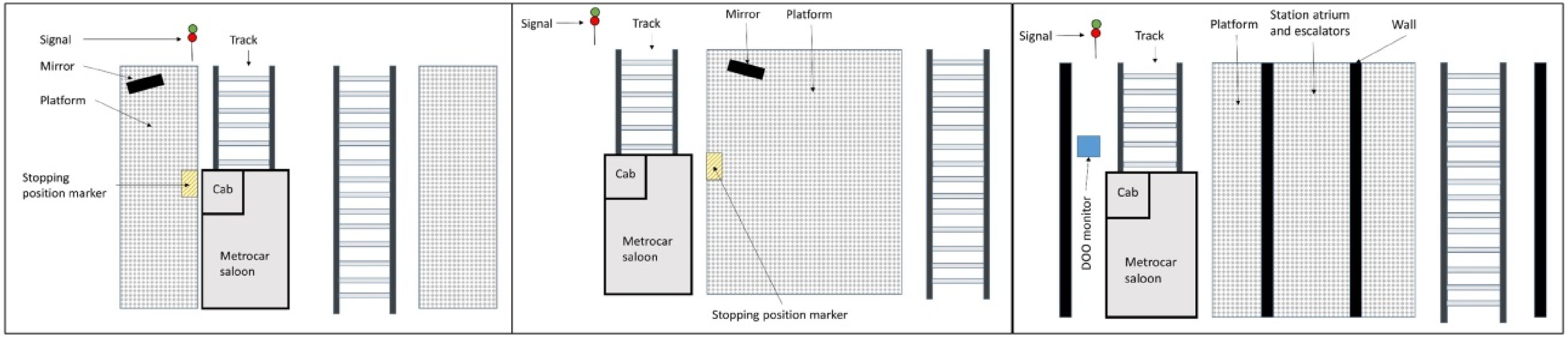


**Figure 8. Three main layouts of Tyne & Wear Metro stations. From left to right: Type 1, Type 2 and Type 3.**

| Station type | Stations |
|---|---|
| **Type 1** | Bank Foot, Bede, Benton, Brockley Whins, Byker, Callerton Parkway, Chichester, Chillingham Rd, Cullercoats, East Boldon, Fawdon, Fellgate, Four Lane Ends, Hadrian Rd, Hebburn, Heworth, Howdon, Ilford Rd, Jarrow, Jesmond, Kingston Park, Longbenton, Manors, Meadow Well, Millfield, Monkseaton, North Shields, Pallion, Palmersville, Park Lane, Percy Main, Regent Centre, Seaburn, Shiremoor, Simonside, South Gosforth, South Hylton (1 platform terminus), South Shields (1 platform terminus), St James (2 platform terminus), St Peters, Stadium of Light, Tyne Dock, Tynemouth, University, Walkergate, Wallsend, Wansbeck Road, West Jesmond, West Monkseaton, Whitley Bay |
| **Type 2** | Airport, Felling, Gateshead Stadium, Northumberland Park, Pelaw, Sunderland |
| **Type 3** | Central Stations, Gateshead, Haymarket, Monument |

**Table 5. Station types by location**

## 4.5 Command and control

Tyne & Wear Metro operates on its own infrastructure, as well as some sections using Network Rail (mainline) infrastructure; thus there is a variety of signalling used. Most of the system has simple 2-aspect signalling, with occasional fixed distants and 3-aspect signals. However, Pelaw to Sunderland route uses Network Rail infrastructure and consequently utilises standard UK mainline 4-aspect signalling, with yellow and double yellow signals. This section is shared with both passenger and freight mainline trains, which operate at various speeds. The depot and turnback locations also have running signals on the ground level – Ground Position Lights (GPL). As of 2016, only GPL at the depot use Light-Emitting Diode (LED) technology.

Tyne & Wear Metro drivers do not benefit from the Automatic Warning System (AWS), or the Train Protection and Warning System (TPWS), available to mainline train drivers at the Network Rail section. A fixed block command & control system is used in Tyne & Wear Metro. The Automatic Train Protection (ATP) system controls overspeeding and SPaDs at certain locations only. The ATP system used is the Indusi system, which is a version of the German mainline railway warning and supervision system Induktives Sicherungssystem. The Indusi system is very coupled with a very efficient metrocar braking system which allows stopping the train from the maximum line speed (80 km/h) in 150 metres. More information on the Tyne & Wear Metro technical specifications can be found in the literature (Howard, 1976; Fenner, 2002; Powell *et al.*, 2014).

## 4.6 Tyne & Wear Metro in context of urban rail and mainline systems

To extrapolate the findings of this thesis to other systems outside of Tyne & Wear Metro, it is necessary to understand the operational differences compared to common urban rail types. Table 6 below provides a comparison between Tyne & Wear Metro and typical urban rail modes. Features of the compared systems are based on the data from Profillidis (2000); (Bonnett, 2005); Vuchic (2007); Powell and Palacín (2015).

| **Parameter** | **Commuter railway** | **Rail rapid transit system** | **Light rail transit** | **Tyne & Wear Metro** |
|---|---|---|---|---|
| **Right of way** | Grade-separated with level crossing (full signal override and gate protection). | Grade-separated | Grade separated with signal priority at level crossings. Street running sometimes. | Grade-separated with level crossings (signal override but no gate protection). A number of uncontrolled pedestrian crossings. |
| **Traction** | Typically diesel or electric (25kV AC) | Electric (1500V DC overhead, 750V DC conductor) | Electric (mainly 750V DC) | Electric (1500V DC) |
| **Station spacing (m)** | 1200 - 4500 | 500 - 2000 | 350-1600 | 400 - 2000 |
| **Maximum speed** | 80-130 km/h | 80-100 km/h | 60-100 km/h | 80 km/h |
| **Number of cars** | 1-10 | Typically up to 10 | 2-4 cars | 2 cars (articulated) |
| **Max Frequency (per hour)** | 10-30 | 20-40 | Up to 60 | 18 |
| **Train Control** | Predominantly manual/signals | Manual/Signals or Automatic | Predominantly manual/visual | Manual/Signals |
| **Signalling** | 4-aspect | 2-aspect (sometimes with repeaters) | Car signalling or 2-aspect | 2-aspect (sometimes with repeaters) and 4-aspect |
| **Personnel** | Driver and a train guard | Driver/attendant | Driver only | Driver only |
| **Cab layout** | Full car width | Full car width | Full/part car width | 2/5 car width |
| **Max acceleration rate ($m/s^2$)** | 0.75 - 1.0 | 1.1 - 1.3 | 1.0 - 1.3 | 1.0 |
| **Emergency deceleration rate ($m/s^2$)** | 0.8 – 1.4 | 1.3 - 1.4 | 2.5 - 3.0 | 2.1 |
| **Train dispatch** | Predominantly train guard assistance | Driver Only Operation (platform monitors or mirrors) | Driver Only Operation (side mirrors or in-cab monitors) | Driver Only Operation (platform monitors or mirrors) |

**Table 6. Tyne & Wear Metro parameters compared to typical urban rail modes**

Tyne & Wear Metro uses a combination of the features typical to different urban rail modes. It uses the signalling common to both rapid transit and commuter railway systems whereas its rolling stock is rather typical to light rail systems. Infrastructure-wise there are similarities with rapid transit systems (station spacing, dispatch methods and traction) but the Tyne & Wear Metro's train control and right of way is closer to a commuter railway. Tyne & Wear Metro had become a melting pot of technological advancements and legacy railway infrastructure when it was built in 1970s. In many ways the system falls into the light rail rapid transit classification as per Vuchic (2007). It operates on the lines typical to rail rapid transit systems but with the rolling stock commonly used for light rail transit. "*This mode is used on lines that require high performance, but do not have the high passenger numbers to justify long trains and large stations*" (Vuchic, 2007, p. 72).

## 4.7 Operational context

According to representatives of the operator, several important changes have been introduced to company culture since the start of the franchise. A "no blame" approach and focus on increased reporting were among those. However, the ATP used in Tyne & Wear Metro cannot notice many incident types so random speed checks and On-Train Monitoring Recorder's (OTMR) log checks occur. With better incident sample at their disposal, the new management started devoting attention to driver-related incident types which do not cause ATP activation, e.g. wrong side door releases. However, SPaDs are still an important issue for the operator. Despite an overall downward trend in those, there are fluctuations present, similarly to national statistics. Interest in underlying reasons for these fluctuations initially provided an opportunity for this research.

Tyne & Wear Metro currently uses a series of state of the art developments from Human Factors field. Incident investigation from a system point of view is ensured by training the investigator to the highest possible level. Non-Technical Skills (NTS) are introduced into driver' training and re-training. The NTS are known to significantly improve overall safety-related performance of train drivers (Russell *et al.*, 2013). Some aspects of system design have been already assessed by the operator, for example signal sighting using tools developed by RSSB. However, these assessments were not comprehensive studies and were mostly based on a trial & error method.

# Chapter 5. Tyne and Wear Metro historic incident data analysis

## 5.1 Introduction

The literature review strongly suggests that design-related PSFs can have important influence on incident propagation. Hence it is necessary to explore incident trends across the investigated system (Tyne & Wear Metro) to address research objective 2. The previous research mostly utilises past incident analysis for this exercise thus utilising similar methodologies allows comparison of the findings. However, the results are assessed in context of the metro system under investigation. For this, the chapter combines quantitative assessment of the incident statistics with the semi-structured interviews and workshops.

## 5.2 Methodology

The methodology for this chapter is designed with the above mentioned objectives in mind but from HF and PSF perspective. Thus it is focusing on incidents where drivers' error was a failure mechanism. Historic incident statistics allows identifying incident "hot spots" with associated design features. Personal and environmental factors are also explored. Inter-dependence and similarities in factors are addressed through correlation analysis. The semi-structured interviews provide additional dimension to the discussion from drivers' point of view.

### *5.2.1 Input data*

The raw data used in this research was provided by the Tyne & Wear Metro operator. 1282 incident reports from the 2011/12 and 2012/13 operating years were made available. Each operating year is divided into 13, four-week long periods, and Table 7 shows the dates for each period of 2012/13.

| Period | Dates |
|---|---|
| **1** | 1 Apr -29 Apr |
| **2** | 30 Apr – 27 May |
| **3** | 28 May – 24 Jun |
| **4** | 25 Jun – 22 Jul |
| **5** | 23 Jul – 19 Aug |
| **6** | 20 Aug – 16 Sep |
| **7** | 17 Sep – 14 Oct |
| **8** | 15 Oct – 11 Nov |
| **9** | 12 Nov – 9 Dec |
| **10** | 10 Dec – 6 Jan |
| **11** | 7 Jan – 3 Feb |
| **12** | 4 Feb – 3 Mar |
| **13** | 4 Mar – 31 Mar |

**Table 7. Dates of the 13 operating periods of 2012/13**

The data is based on the entries contained in an incident reporting system. Each incident is logged with a brief description, date, time and location, as well as being allocated an incident type. Such incident types are grouped into larger categories, e.g. “technical domain incidents”. The data is unanimous thus anthropometric data, experience and age cannot be factored in. The current reporting system is focused on operational incidents, particularly technical faults. To perform the analysis, a new categorisation of the incident reports was required, as well as selecting the incident types to be retained. The categorisation, incident types used, and descriptions are summarised in Table 8. The categorisation is focused on driver-related incidents, but some of the technical and operational domain incidents are also retained. Incident reports that did not fit any of the categories are excluded from the study. These include mainly incidents associated with pedestrians or vehicles trespassing at level crossings. Statistics on dispatch equipment faults is only available for 2012/13 operational year.

| Category | Incident type | Description |
|---|---|---|
| **SPaD incidents** | Category A SPaD | Category A Signal Passed at Danger (SPaD) occurs when a driver passes a signal displaying a stop aspect, without a permission due to own error |
| | Other SPaDs | Occurs when a driver passes a signal displaying a stop aspect, without permission, due to a technical fault |
| **Driver-related incidents** | Overspeeding | A driver exceeds the maximum permitted line speed |
| | Failure to call | A driver skips a scheduled stop, without instruction |
| | Station overrun | Occurs when a train stops beyond the platform end or beyond a point allowing to dispatch train safely |
| | Passenger entrapment | When a passenger is trapped by train doors during door closing procedure |
| | Wrong side doors activation | A driver releases a wrong set of doors |
| | Wrong route | A driver sets an incorrect route code that later affects passenger information and route setting. It is also possible for the system to set points incorrectly due to technical fault. A driver who does not notice the issue, and takes the wrong route, is still at fault. |
| **Technical domain incidents** | Signal faults | Technical faults associated with signalling equipment |
| | Dispatch equipment faults | Faults of Driver Only Operation (DOO) equipment, e.g. misting up or out-of-focus monitors |
| | Trainfault ATP | Technical faults of train-borne ATP equipment |
| | Trackfault ATP | Technical faults of lineside ATP equipment |
| **Operational incidents** | Doors obstruction | Passengers restricting train doors from closing |
| | Passenger overcarried | Passengers being left on a train when a train is taken out of passenger service |
| **Interface incidents** | LRA incidents | Low Rail Adhesion (LRA) conditions that did not lead to a more serious incident |
| | Foliage foul of infrastructure | Foliage or vegetation creating risks for safe use of Tyne & Wear Metro infrastructure, usually DOO equipment |

**Table 8. Categorisation of the incidents used in the study and their descriptions**

It can be seen that some of the incident types are similar in their definitions. For example, passenger entrapments and doors obstruction incidents are typically very similar in their nature. Almost all door obstructions are reported by drivers due to causing some technical faults (doors cycling) or delays. If there was a risk of trap & drag accident or an injury, e.g. doors did not cycle, it is reported as passenger entrapment. However, the final category is allocated by a safety manager based on driver’s report and other support tools, e.g. CCTV footage.

Usually incidents are sub-divided into, as per IOSH categorisation (Institution of Occupational Safety and Health, 2016), into accidents and near-misses. The main difference between those two types is whether any harm was caused or not respectively. However, due to lack of details in the incident reports, it was decided to treat all of the entries just as incidents. Even if those all are near misses, it indicates significant risks and should be treated as precursors.

Three separate data sets based on time, period, and location were created, after assessing the raw data for consistency. This was required due to incident causal mechanisms related to time, date and location usually fall into different PSFs groups. Time and date-related factors are associated with lighting and climatic conditions, patronage numbers and demographics, and seasonality. Furthermore, time-related factors also include individual driver factors, e.g. circadian rhythms. On the other hand, the location-related data set allows for an investigation of system design features influencing Tyne & Wear metro drivers’ performance. Driver-related incidents are understood as a combination of different factors affecting a driver. Such a combination of factors, happening in a certain order, is capable of bypassing the safety mechanisms of a system.

Several incident types have a relatively small number of cases, e.g. station overrun, and failure to call incidents. Locations are understood as an approach station. Most of the driver-related incidents in this analysis are associated with station-based duties, thus approach station information is considered to be sufficient. All timeframes for the time-based data set are one hour long, e.g. 07:00:00 to 07:59:59, with the exception of 01:00:00 to 04:59:59, when only maintenance trains operate on the system. The Tyne & Wear Metro passenger flow has two peaks: between 8 and 10 am, and from 4 to 6 pm. To cope with the increased passenger numbers, the number of trains in the network is at its highest during these hours.

As discussed in Chapter 1, category A SPaDs have become a certain benchmark in the railway industry. Hence the older data for 2009/10 and 2010/11 operational years for this incident type was deemed of good quality and reliability to expand descriptive statistics. Concerns for data quality and reliability of it prior to 2011/12 operational year were expressed by some Tyne & Wear Metro managers in a preliminary interview. These concerns are also supported by significant difference in incident records between 2011/12 and 2012/13, where the number of increases by 43% in 2012/13 operational year.

All statistical manipulations are performed in IBM Statistical Package for Social Sciences (SPSS), version 22.

### *5.2.2 Normality of the data and other considerations*

All variables in the created data sets are checked for normality as this affects choice of statistical analysis tools. Shapiro-Wilk and Smirnov-Kalmogorov tests are chosen for these tests. Sample size defines which test result to use in the assessment of normality. If a tested sample size is less than 50 then Shapiro-Wilk test is performed, otherwise output from Smirnov – Kalmogorov test has to be considered (Marques de Sa, 2007).

The count nature of the data in the samples leads to the zero-inflated data issue. The Cambridge Dictionary of Statistics (Gibson, 2016) defines the zero-inflated data as "*count data with excess zeros*" (p.467). Tu (2006) emphasises that "*the probability mass at the point zero exceeds that allowed under a standard parametric family of distributions*" (p.1). It is important to mention that all the zeros in the samples are genuine zeros. There is a range of opinions on what to do with this kind of data, from choosing methods that fit data to transformation of variables to normally distributed ones (Vives and Losilla, 2006).

Conventional transformation techniques, e.g. square root or log transformations, cannot be applied to zero-inflated samples. Vives and Losilla (2006) indicated that the distributional character of a count variable has to be respected. They quote Gardner *et al.* (1995) who suggested that the statistical methods have to be fitted to data not vice versa. Furthermore, transformations make results significantly harder to interpret (Vives and Losilla, 2006). For example, using Box-Cox transformation (Box and Cox, 1964) yields different transformation formula for each of the variables. This raises the question about back-transformation of relationships found between two variables with different formulas applied. Huson (2007) showed that standard correlation techniques are adequate for analysis of zero-inflated data. However, he did not test samples with more than 30% of zeros so cannot be applicable to some of the variables in this chapter, e.g. station overruns. Simple transformation utilised by Spearman Rank Correlation (SRC), i.e. converting data into ranks, yields similar results to more sophisticated transformations along with Pearson Product-Moment Correlation (PPMC) (Hazarika, 2013).

As the majority of statistical manipulations performed in this chapter are correlation tests, it is assumed that the data does not require transformation because conventional methods provide good validity.

### *5.2.3 Frequency distribution analysis*

The frequency distribution of each incident type is studied using bar charts, in order to identify peak times, seasonally inflated incident rates, and the worst performing locations. The current approach to incident analysis often focuses only on statistics from the entire network, whereas there is also a need to analyse local trends and risks (Bearfield *et al.*, 2013). The four worst performing stations for certain driver-related incident types have been selected to carry-out in-depth examination of the potential factors inducing incident propagation at these specific locations.

A preliminary analysis of descriptive statistics led to several hypotheses, which required additional data collection.

### *5.2.4 Additional indicator data*

Additional data was collected from open-access sources, e.g. Google Maps, and the Tyne & Wear Metro operator in order to check how the findings from previous research apply to urban rail. The four key areas have been assessed i.e. i) passenger composition, ii) driver shift patterns, iii) tunnel exit/entrance, iv) distances between the stations.

Passenger composition is not uniform throughout a typical day and does not fully match with number of trains in the system. Certain age groups tend to use Tyne & Wear Metro more often during certain time periods. Specifically, school children use Tyne & Wear Metro the most from 7.30 to 9 am, and 3 to 5 pm. Elderly people tend to use Tyne & Wear Metro between 9.30 am (when concessionary travel time starts) and the evening peak. These 2 demographics are potentially at higher risk to be involved in certain incidents, e.g. passenger entrapments. According to RSSB *et al.* (2006) presence of mobility equipment, elderly or very young passengers had been a contributory factor in approximately 20% of the mainline PTI-related incidents between 2001 and 2005. With a demographics breakdown of station patronage levels not available, it is assumed that presence of schools and hospitals within 500m radius from a station is an indicator of increased usage of that facility by the above mentioned population groups.

Research from mainline railways shows increased risks of incidents 2 hours into a driving portion due to presence of vibrations negatively affecting arousal levels (Keun Sang and Ohkubo, 1994; Yang *et al.*, 2012). Similar observations were found in other safety-critical industries (Folkard, 1997; Macdonald *et al.*, 1997). Data on the hourly number of trains in the system along with data on hourly driver sign-on was sourced to check for this.

Previous research suggests that a drastic change in environment at tunnel exit/entrance should positively affect arousal levels (Yang *et al.*, 2012). 24 locations associated with tunnel entrance/exit (underbridges are not included) have been explored to provide comparison with the rest of the network. Monotonous environments are known to cause rapid sensory deprivation, at least among car drivers (Thiffault and Bergeron, 2003; Williamson *et al.*, 2011). Assuming that negative effects of such deprivation increase with time travelled, distances between the stations were collected.

### 5.2.5 Bivariate correlation analysis

A bivariate correlation analysis has been selected to interrogate the data in order to:

- explore the existence of correlation between statistics, for two operating years, for each driver-related incident type, in order to check consistency year-on-year;
- uncover potential associations between different incident types in all the data sets;
- explore possible associations between the additional indicators and driver-related incident types in the respective data sets.

Consistency year-on-year in a driver-related incident type should suggest a connection between the incident type and certain locations, times or periods. Associations found between the incident types indicate potential presence of common causal factors. Exploring similarities in causal factors can help with understanding of the inter-dependency of PSFs in metro systems. It is important to note that a correlation found does not always suggest causation. In many cases the relationships will be spurious correlations, meaning that there are common factors affecting both variables.

PPMC and SRC tests are carried out depending on normality, respectively when both variables are normally distributed and when at least one is not normally distributed. Use of both methods is required as PPMC is sensitive to non-normality (Kowalski, 1972). However, SRC does not find statistically significant associations for monotonic associations, e.g. when a scatter plot produces a U-shaped graph. Hence the scatter plots were checked for all relationships tested with the SRC but no monotonic associations were found.

### *5.2.6 Category A SPaD review for situational awareness*

Direct observations carried out in the Tyne & Wear Metro control room in 2015 suggested correlation between disruptions in the system and category A SPaDs. This theory elaborates on the previous incident investigations, e.g. SPaD in New South Wales in November 2014 (Australian Transport Safety Bureau, 2015). Train drivers sometimes fail to notice red aspects because they do not expect to see those. On the other hand, serious disruptions usually lead to trains being re-routed or turned back in unusual locations thus causing more red aspects. There are no means of communicating an alert about a disruption ahead to a large number of drivers so a task of updating each individual driver sits with system desk controllers (SDCs). In events of big disruptions, warning a driver about an unexpected signal at danger might not be on SDCs' list of priorities. Decrease in situational awareness can become a major risk factor as drivers lose ability to predict future status of the system (Parkes and Hooijmeijer, 2001). This is found to contribute to poor driving performance in automotive industry. Moreover, disruptions in the system are highly associated with increased workload for operators (Balfe *et al.*, 2015).

An additional Category A SPaD assessment was conducted to check the above hypothesis. Category A SPaD investigations are used to determine whether drivers could be "caught by surprise" by one of the disruptions prior to SPaD event. Total of 32 investigations were studied for period between July 2013 and November 2015 using a train movement replay system available in Tyne & Wear Metro. Consulting with experienced SDCs and incident investigators, the incidents were assessed in terms of potential contributing factors. The sample of incidents is limited by the train movement replay system's ability to provide archive information.

### *5.2.7 Workshop with DB Regio Tyne & Wear*

A workshop with the operator's key safety staff and drivers' representatives was organised in 2014 to validate preliminary findings of the statistical analysis. This semi-structured workshop was run by the author and allowed complementing the findings with spatial and timely context often only available to people with long-term experience in the system. Among those attending were Head of Safety & Compliance, Head of Operations, Traincrew Manager and a representative of Tyne & Wear Metro drivers. Most of the attendees have some metro driving experience and are RSSB-certified incident investigators. The workshop started with a presentation by the author on results of the data analysis. In the end of the presentation pre-selected discussion points were given. The pre-selected points were based on the results of the statistical analysis which, in author's opinion, required explanation from the participants. Those include differences in depot layouts, differences in SPaDs and safety culture, evening peaks in overspeeding, station overruns at Fellgate. The discussion was relatively open with only a little moderation required in order to steer the direction of it.

### *5.2.8 The semi-structured interviews with Tyne & Wear Metro drivers*

Monthly visits were arranged to the Tyne & Wear control centre, where drivers usually spend their brakes between driving portions. This presented the author with opportunities to explore drivers' views on findings and developed theories as the research progressed through different phases. More than 60 semi-structured interviews were carried out. At least three different drivers were interviewed at the Tyne & Wear Metro control centre each time. A typical interview lasted for 10 to 15 minutes and covered 3 questions, which were defined as the data analysis progressed.

With the interviews being a continuous process spanning the entire research, the research questions addressed are not exclusive to Chapter 5. The interview results are incorporated in discussion sections throughout the thesis.

## 5.3 Results

The methodology described in Section 5.2 has been applied to analyse the data sets produced based on the Tyne & Wear Metro incident statistics.

### *5.3.1 Consistency of the driver-related incidents*

Table 9 below shows the consistency in distribution of driver-related incidents over the two operating years. The three data sets are presented in different columns. Most of the driver-related incidents are localised to certain parts of the network. Only two out of seven incident types are consistent in terms of times. Category A SPaDs are the only driver-related incident type to show year-on-year correlation in the context of periods. However, as the correlation coefficient is negative, the distribution of this incident type is not consistent.

| Incident type/Sample type | Date-based | Time-based | Location-based |
|---|---|---|---|
| **Category A SPaDs** | -0.565* | -0.123 | 0.283* |
| **Overspeeding** | -0.013 | 0.244 | 0.400** |
| **Failure to call** | 0 | 0 | 0 |
| **Station overrun** | 0.459 | 0.221 | 0.281* |
| **Passenger entrapment** | -0.065 | 0.601** | 0.353** |
| **Wrong side doors activation** | -0.012 | 0.427* | 0.543** |
| **Wrong route** | -0.226 | 0.395 | 0.374** |

**Table 9. Correlation of driver-related incidents, year-on-year; * *p*-value<0.05; ***p*-value<0.01**

### *5.3.2 Driver-related incidents*

Table 10 shows a composition of the driver-related incidents in Tyne & Wear Metro. For the two years under investigation, overspeeding and wrong side doors incidents are the most encountered types, with 26.3% and 26.6% respectively. The incidents associated with station procedures (wrong side doors activation, passenger entrapment) account for 46.6% of the entire data set.

| Incident type | 2011/12 | 2012/13 | For 2 years |
|---|---|---|---|
| **Category A SPaD** | 6.3% | 10.4% | 8.4% |
| **Overspeeding** | 27.5% | 25% | 26.3% |
| **Failure to call** | 2.8% | 0% | 1.2% |
| **Station overrun** | 6.3% | 4.8% | 5.6% |
| **Passenger entrapment** | 13.4% | 26.4% | 20.0% |
| **Wrong side doors activation** | 36.7% | 16.7% | 26.6% |
| **Wrong route** | 7.0% | 16.7% | 11.9% |
| **Total** | 100% | 100% | 100% |

**Table 10. Proportion of driver-related incidents by year and total.**

When the totality of the driver-related incidents is considered, there are 3 pronounced peaks in incident propagation (Figure 9). The morning peak in incidents fully coincides with the morning peak in passenger levels (8 to 10 am). The midday peak can be observed between 1 and 3 pm. The evening peak also partially coincides with the evening peak in passenger level (4 to 6 pm) but lasts until 7pm.

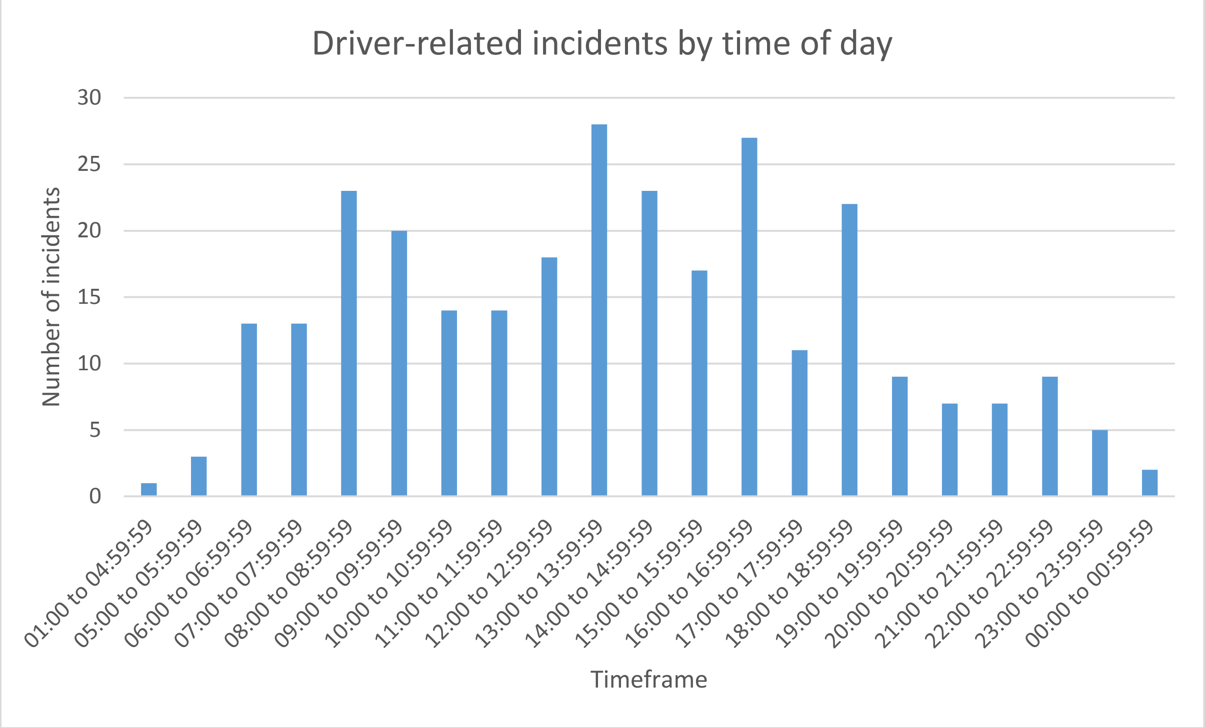


**Figure 9. Composition of driver-related incidents by time of occurrence (2011/12 and 2012/13 operational years)**

Figure 10 shows that there are more driver-related incidents occurring in the second half of the operational year. The rise starts in period 8 (mid-October) and continues through the winter. This time is associated with increased passenger loadings due to Christmas shoppers and mostly night time driving. Period 6 also demonstrates high incident numbers. The biggest half-marathon in the world, Great North Run, is always held during period 6, putting additional pressures on the system.

Lack of consistency for variables in date-based and time-based samples shows that seasonal and time of the day distributions might differ from year to year. The count data for all of the driver-related incident types is summarised in appendixes A to C. The statistics for four worst performing locations for overspeeding, station overruns, passenger entrapments and wrong side door activations can be found in appendixes D to G. Category A SPaDs are summarised for four consecutive operational years thus providing more reliable sample. Only three stations have ‘failure to call’ occurrences in the two years under investigation. For wrong route incidents the stations account for 55.9% of the cases with fourth worst station only contributing to 6%. Table 11 provides a summary of the findings from the frequency distribution analysis.

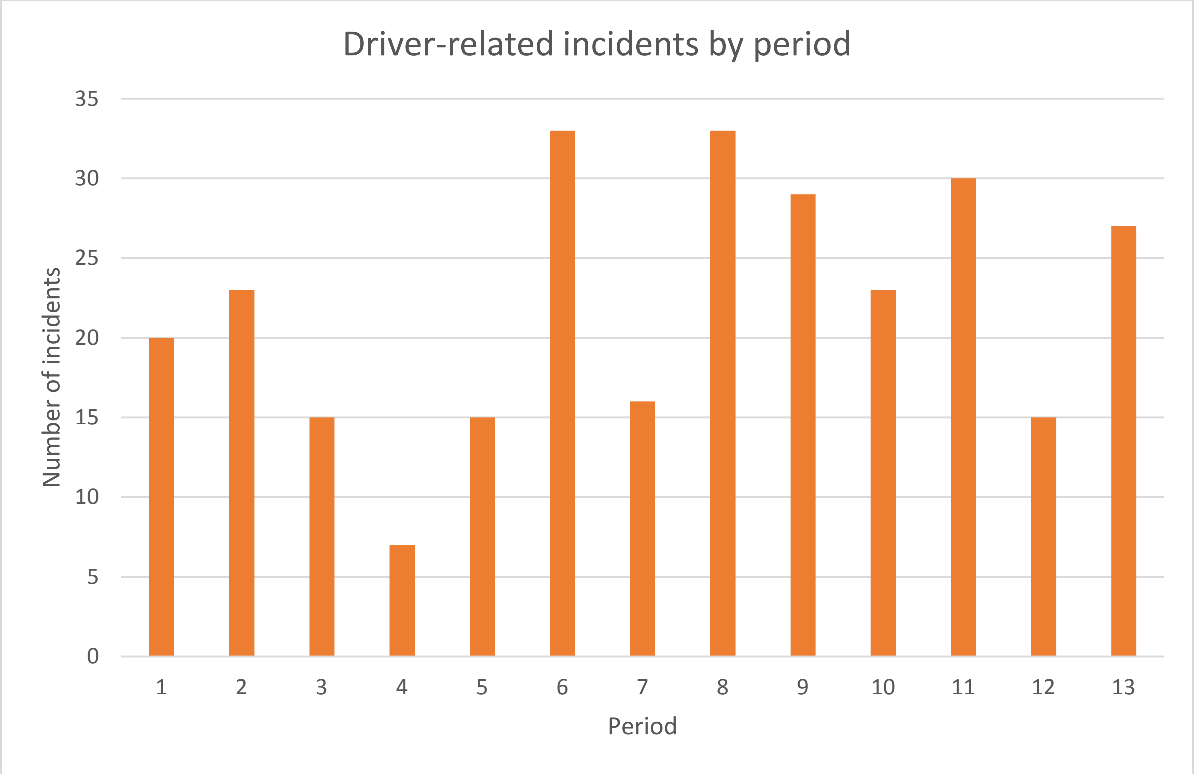


**Figure 10. Composition of driver-related incidents by period of occurrence (2011/12 and 2012/13 operational years)**

Category A SPaDs peak around July and February with relatively lower incident levels in-between those. December and mid-August to mid-September also have inflated incident levels compared to average values. Two daily peaks are distinguishable between 8 and 12 am, and 4 to 7 pm. The worst performing locations are Pelaw (Type 2), Regent Centre (Type 1), Jarrow (type 1), Airport (Type 2 terminus), and the depot. The types used in the description are the station types presented in the Tyne & Wear Metro overview (Section 4.2). All of the SPaDs happened at Jarrow in 2009 (Type 1).

Overspeeding incidents have different distribution throughout the two years. The most incidents are observed from April to May, and from mid-October to mid-November. Apart from the 6 to 7 am timeslot, a steady growth can be seen between 9 am and 5 pm. The worst performing stations are Airport (Type 2 terminus), Sunderland (Type 3), Manors (Type 1), South Hylton (Type 1 terminus with 1 platform). At these locations incidents mostly happened between periods 1 and 3, and in period 6.  Time distribution of the incidents at four worst stations is similar to full sample, with the 3 to 5 pm peak being more pronounced.

There are only four cases of ‘failures to call’, all in 2011/12. Those happened in periods 2, 7, 8 and 9 automatically making those peak periods. Two of the incidents happened between 1 and 2 pm. Other two are either early morning or late night occurrences. As it was mentioned before, the incidents were localised to only three areas – Meadow Well (Type 1), Northumberland Park (Type 2) and University (Type 1).

Station overruns are most often observed in periods 8 and 9 and in winter months after New Year. Three peaks can be seen: 8 to 9 am, 11 am to 12 pm, and 3 to 4 pm. The worst performing locations are East Boldon (Type 1), Hadrian Road (Type 1), Central Station (Type 3) and Fellgate (Type 1). At these stations the peaks in station overruns are periods 9 and 11, but no daily peak times. However, bigger concentration of the station overruns can be seen in the morning hours. Considering a strong association of overruns with failures to stop (the former is potentially a precursor of the latter), it is important to note that Meadow Well and Northumberland Park have one overrun incident each.

The periodic peaks for station overruns match periods when the system is affected by LRA issues. However, Fellgate is the only station out of the four worst performing locations where the overruns are observed outside of the normal LRA periods. Fellgate and East Boldon deviate from the standard Type 1 design due to a running signal being located considerably further from platform edges than normally.

There are significantly more passenger entrapments reported in 2012/13 operational year. This incident type becomes more pronounced between late Autumn and early

Spring but not in period 12. Daily peaks occur right after midday and inflated incident rates can be observed during the peak hours too. The majority of the worst performing locations are Type 3 stations with only Monkseaton being Type 1 station. When only the four worst performing locations are considered, the daily peaks are the same as for the entire sample. On the contrary, the monthly data for the locations suggests more spread out distribution of these incidents with only period 11 having a considerable peak. Period 11 usually includes four weeks after New Year.

Wrong side door activation incidents tend to occur more in the second part of an operational year with a peak in period 6. However, most of the incidents contributing to the peak occurred in 2011/12. For the other year under investigation the incidents are more spread out but an increase can be seen in January-March. Similarly to passenger entrapments, three peaks are observed. Two of those coincide with maximum passenger flows and the third one (the biggest) is between 12 and 3 pm. The four worst performing stations for this incident type are Haymarket (Type 3), West Jesmond (Type 1), St James (Type 3 terminus) and Northumberland Park (Type 2). All of these stations but West Jesmond involve approach routines with altering platform sides. For example, approaching Haymarket (57% of all the incidents at these locations) from Jesmond means that platform side changes from right to left. St James being Type 3 terminus station means that the drivers can arrive to any of the two platforms. However, most of the incident at Haymarket happened at platform 2 which does not involve platform side change as trains arrive from another Type 3 station. When only the four worst performing stations are taken into account, the monthly distribution stays the same. On the contrary, the morning peak becomes more pronounced for the four stations but there is no evening peak.

Even if Haymarket is considered as an outlier, 15 out of 25 stations with at least one incident have at least one of the approaches where a platform change occurs.

Wrong route incidents, due to their nature, are confined to locations where a choice of routes is present. Pelaw and South Gosforth are locations of two main junctions in the system, whereas Regent Centre provides several routes into the depot. All of these locations also have turnback facilities, which are actively used for peak services. This also affects the usual occurrence time which coincides with the passenger flow peaks and presence of additional short peak services using the turnback facilities. Periods 6 to 8, 10 and 13 have the highest wrong incident rates. However, these peaks are caused significant rise in reporting in 2012/13.

### *5.3.3 Correlation analysis*

The results of the correlation analysis are demonstrated in tables 1 to 3 in Appendix H. In total, 60 associations were found between incident types. 49 of those involved driver-related incidents. In terms of such incidents, 53% of the associations were discovered in the location-based data set and only 14% in the period-based data set. In the time-based sample, Category A SPaDs, passenger entrapments and wrong side doors activations have the highest number of associations (4). Majority of associations found in that sample have medium to low strength. The majority of location-based associations are low in strength, with four statistically significant negative correlation coefficients found, mostly with passenger entrapment incidents. Category A SPaDs are involved in almost 25% of the correlations with driver-related incidents found in the location-based data set. Table 11 provides a summary of the findings from correlation analysis.

|  | Peaks |  | Associations (correlation coefficient) |  |  |
|---|---|---|---|---|---|
| Incident type | Peak period(s) | Peak time(s) | By period | By time | By location |
| Category A SPaD (4 years of data) | 4, 12 | 4-7 pm;<br>8-12 am | None | Station overrun (.498*);<br>Passenger entrapment (.574**);<br>Wrong side doors (.601**);<br>Trainfault ATP (.524*); | Overspeed (.256*), Wrong route (.334**), Signal faults (.269*), Trainfault ATP (.301*), Trackfault ATP (.260*), Passenger overcarried (.526**); |
| Overspeeding | 1-2, 8 | 2-5 pm | None | Passenger entrapment (.772*);<br>Wrong side doors (.453*);<br>Doors obstruction (.472*); | Category A SPaD (.256*);<br>Trackfault ATP (.307*); |
| Failure to call | 2, 7-9 | 1-2 pm | Trackfault ATP (.600*);<br>Dispatch equipment (.566*); | None | Station overrun (.282*); |
| Station overrun | 8-12 | 8-9 am;<br>11 am – 12 pm,<br>3-4pm | Dispatch equipment (.666*);<br>LRA (.718*); | Category A SPaD (.498*);<br>Foliage fouls (.580**); | Failure to call (.282*), Passenger entrapment (.273*), Foliage fouls (-.407**); |
| Passenger entrapment | 8-11, 13 | 12-3 pm | Doors obstruction (.578*);<br>Dispatch equipment (.633*); | Category A SPaD (.574*);<br>Overspeeding (.772**);<br>Wrong side doors (.696**);<br>Doors obstruction (.606**) | Station overrun (.273*), Doors obstruction (.280*), Foliage fouls (-.377**), LRA (-.321*), Passenger overcarried (-.257*); |
| Wrong side doors | 6, 8-9, 11-13 | 1-3 pm;<br>7-10 am;<br>4-5 pm | None | Category A SPaDs (.601**);<br>Overspeeding (.453*);<br>Passenger entrapment (.696**);<br>Foliage fouls (.606**); | Trackfault ATP (.285*); |
| Wrong route | 6-8, 10, 13 | 8-9 am;<br>6-7 pm | Dispatch equipment (.579*); | LRA (.450*);<br>Passenger overcarried (.441*); | Category A SPaD (.334**), Signal faults (.293*), Foliage fouls (.398**), LRA (.261*), Passenger overcarried (.475**) |

**Table 11. Summary of the results. * p-value < 0.05 (2-tailed); **p-value < 0.01 (2-tailed).**

### 5.3.4 *Correlation with additional indicators*

Table 12 below displays the correlations found between the additional data collected and driver-related incident types. Data on 'distance between stations' and 'schools and hospitals nearby' were analysed for correlations with driver-related incidents from the location-based data set. The remaining additional data was checked for correlation with driver-related incidents from the time-based data set.

| | Distance between the stations | NR of hospitals and schools nearby | Number of trains in the network per hr | Number of drivers starting their shift per hr | Number of drivers who started their shift ~2 hrs before (per hr) |
|---|---|---|---|---|---|
| Category A SPaDs | 0.017 | 0.008 | 0.632** | 0.068 | 0.535* |
| Overspeeding | -0.119 | -0.222 | 0.528* | 0.304 | 0.435* |
| Overspeeding (4 worst performing) | N/A | N/A | 0.408 | 0.292 | 0.274 |
| Failure to call | -0.111 | 0.096 | 0.115 | 0.176 | -0.031 |
| Station overrun | 0.094 | -0.026 | 0.360 | 0.191 | 0.438* |
| Station overrun (4 worst performing) | N/A | N/A | 0.168 | 0.016 | 0.240 |
| Passenger entrapment | -0.224 | 0.086 | 0.594** | 0.086 | 0.342 |
| Passenger entrapment (4 worst performing) | N/A | N/A | 0.330 | -0.023 | 0.116 |
| Wrong side doors activations | -0.213 | -0.195 | 0.794** | 0.224 | 0.580** |
| Wrong side doors activations (4 worst performing) | N/A | N/A | 0.528* | 0.247 | 0.414 |
| Wrong route | 0.167 | 0.166 | 0.130 | 0.012 | 0.491* |

**Table 12. Associations with additional indicators. * p-value < 0.05; ** p-value <0.01**

24 locations associated with tunnel exit/entrance on the approach were identified and explored, demonstrating 10% higher incident levels compared to the rest of the Tyne & Wear Metro network.

### 5.3.5 *Additional Category A SPaD analysis*

The closed nature of the system means that Tyne & Wear drivers, especially experienced ones, can develop very good knowledge of each driving portion. This knowledge also involves ability to forecast any red signal aspects because of known timetabled conflicting movements. Out of 32 category A SPaDs investigated 16 were found to be associated with red aspects, which not encounter routinely by drivers. Some examples of what might cause such red aspect include:

- Movements of engineering locomotives and rail head treatment trains;
- Trains held at stations waiting for an ambulance or due to an operational incident;
- Signal and infrastructure failures;
- Movements of other train operating companies at Network Rail infrastructure;
- Train failures;

In 44% of incidents identified as disruption-related, the drivers do not have any means of prior warning which would advise them about an unexpected signal aspect. This happens due to 2-aspect signalling and no repeaters in certain areas. Network Rail part of the system, with its 4-aspect signalling, has only 4.1% category A SPaDs even though it accounts for 19% of track-kilometres. However, even more disruption-related SPaDs happened where advanced warning through the signals is available. Hence significant contribution of slips and attention lapses has been pointed out by the investigators. For other 50% of category A SPaDs, lack of experience, attention lapses and distractions were identified as main failure mechanisms.

### 5.4 Discussion

The year on year increase in the driver-related incidents is relatively small compared to 43% rise in overall number of incidents reported. Such rise in the incident reporting without a considerable rise in the driver-related incidents demonstrates improvements in a safety culture (Hale *et al.*, 2010) as well as positive results from Tyne & Wear Metro operator’s efforts to increase reporting rates. After taking the Tyne & Wear Metro concession in 2010, the new operator started gradually focusing more on other driver-related incidents, which were previously overlooked due to prioritising of SPaDs. Composition of different driver-related incident types highlights major risks associated with the PTI. According to Basacik *et al.* (2009) the DOO systems carry the second biggest amount of risk to passengers among the different dispatch methods.

The initiatives introduced by the operator can be tracked in composition of driver-related incidents (Table 10). For example, category A SPaDs, which is the only incident type that can be tracked with 100% accuracy, went up after a campaign to reduce those. The drivers complained about an overload of SPaD-related information at the work place. As soon as this information was taken down, the incident rates reduced. The positive behaviour approach of such campaigns can cause negative effects, e.g. stress, anxiety (due to assumed peer pressure) and increased responsibility, as shown by Karowich *et al.* (2009). This also highlights importance of top-down initiatives for better reporting meaning that the data (period-based) in this study should be treated with caution due to being collected during various campaigns.

Lack of consistency in the period-based sample can be a sign of low importance of season-related factors. Even category A SPaDs, deemed to have the best data quality, show reverse correlation between the two years. The driver-related incident composition demonstrates that almost 50% of these incidents involve station procedures, whereas SPaDs only account for less than 10%. With metro drivers encountering a lot of stations stops per hour, it means that for a long time research and operational performance focus had not been aligned with the actual risk profile.

The findings from the consistency analysis (Table 9) confirm importance of design-related PSFs in incident propagation. Even a marginal localisation of each incident type provides good opportunity of assessing these locations for design features, which might have adverse effect on drivers' performance. Moreover, the location-based sample should not be affected by various management initiatives (in theory), similarly to the time-based sample. Consistencies of the daily distributions of the station-related incidents corroborate the effects of patronage levels, crowding and distractions associated with passengers. Moreover, organisational factors, e.g. workability of a timetable (ease to stay on time), are relevant.

### *5.4.1 Category A SPaDs*

The descriptive statistics for the period between 2009 and 2013 shows similar trends to those discovered in the additional category A SPaD analysis (Section 5.3.5). The majority of the driver-related SPaDs tend to happen in locations where conflicting movements can occur. All of the worst performing locations fall into this category. The conflicting movements are often caused by presence of turnback facilities and sidings at these locations. Hence the association with passenger overcarried and wrong route events. One of the unique features of the sidings and depot is presence of ground position lights. As the name suggests, these signals are located on the ground level and are often combined with direction indicators, all of which differ from the arrangements elsewhere in the system. Furthermore, due to the metrocar design visibility of such signals is obstructed from short distances or requires stand-up driving. As some of the SPaDs in these locations happened in sidings, the GPL are potentially an important causal factor.

High proportion of the disruption related SPaDs means that drivers had encountered signals they have not expected to see. As drivers do not receive the most up to date information about the state of the system, their SA drops. Moreover, prolonged stops (waiting for a signal to clear) cause disengagement from a driving task thus increasing risk of SPaDs due to lower SA (Naweed and Rainbird, 2014). Madigan *et al.* (2016) discovered that events not encountered in normal operations are important contributory factor in rail incidents, especially in SPaDs. Repetitive experience can form an expectation bias and cause loss of the SA (O'Connell *et al.*, 2015). The semi-structured interviews revealed that more experienced drivers know how to interpret different signs to build up their situation awareness. For example, not meeting a train in opposite direction in a usual spot advices about a disruption further ahead. Dray *et al.* (1999) also found that experience, or lack of it, can be the PSF for metro drivers.

Research from the automotive industry does not show relationship between experience and SA (Underwood *et al.*, 2013), but it does not take into account repetitiveness of the task in timetabled driving. The problem is acknowledged in the railway industry with work being done on improving drivers’ SA by using emerging technologies, e.g. DAS (Roth *et al.*, 2006; Young and Grenier, 2009; Tschirner *et al.*, 2013). Apart from being low on priority list of SDCs, the reliance on analogue radio system can limit communication capabilities (Roth *et al.*, 2006). One of the solutions explored by Tyne & Wear Metro is linking the in-cab DAS with the customer services twitter account to provide live alerts to Tyne & Wear Metro drivers too.

Associations found with locations of different ATP and signal faults. Even though the semi-structured interviews and workshops did not reveal any associations of these incidents with drivers’ performance, there is a bigger concentration of signals and ATP equipment at stations with conflicting movements. Hence there is an elevated risk of SPaDs but also higher probability of a technical fault.

Tyne & Wear Metro owned infrastructure has significantly inflated category A SPaD rates compared to Network Rail infrastructure. The main difference between two parts of the system is use of different design signalling. Tyne & Wear Metro drivers receive reduced advanced warning on the Metro infrastructure, which is the busiest part of the system. This is not a direct failure mechanism but PSF increasing SPaD risk. It contradicts findings from mainline railways, where Li (2004) demonstrated that 4-aspect signals have more multi-SPaD occurrences. However, these findings do not account for levels of service and frequency of red aspects encountered. In mainline railways 2-aspect and 3-aspect signals are used in less busy systems than 4-aspect counterparts. On the contrary, the Network Rail part of Tyne & Wear Metro does not have so many conflicting movements thus affecting the incident propagation. Such relationship is also supported by category A SPaD correlation with number of trains in the network (see Table 12). The importance of fatigue and arousal levels is demonstrated by statistically significant correlation with proportion of drivers who have been on shift for two hours.

As for non-disruption situations, the causal factors identified in Section 5.3.5 are similar to the findings presented by Li (2004) from mainline railways but with less emphasis on immediate physical environment PSFs. The daily profile of category A SPaDs occurrence in Tyne & Wear Metro is similar to mainline railways except for a night peak, which is usually attributed to freight trains. According to Li (2004) these peaks are consistent with circadian patterns and thus support importance of personal factors. The discussions facilitated during the workshop revealed that many of the SPaDs are Start Against Signal SPaDs (SASSPaD), although specific locations were not provided. This does not allow making any assumptions on effects of specific design features but allows exploring other PSFs. Even though DOO dispatch eliminates risks of poor communication between a driver and dispatcher in SASSPaD propagation, it increases drivers' workload (Basacik *et al.*, 2009). Monitoring of the platform-train interface is a task which competes with signal checks on departure even though a correct sequence is established in the rulebook. The daily incident peaks coinciding with the maximum passenger flow times suggest a more demanding PTI monitoring task which can in turn distract from the signal checking task.

Frequency distribution of category A SPaDs in mainline railways is also similar to Tyne & Wear Metro but for February peak. The peak in summer is assumed to be caused by heat, pollen levels, sun height, foliage, operational factors, sociological factors and individual factors (Li, 2004). Assuming direct relationship of these factors with human performance, there must be summer peaks in other driver-related incidents but none are observed. Hence it is possible to shortlist only seasonal PSFs directly affecting interaction with signals, e.g. foliage obstruction, sun height. On the other hand, most of the driver-related incidents in Tyne & Wear Metro have a peak in period 11 onwards. In the semi-structured interviews, the drivers noted that lower number of daylight hours cause fatigue and drowsiness, whereas sociological factors are more acute due to the “financial pressures imposed by Christmas”.

### *5.4.2 Overspeeding*

According to Monk *et al.* (2015), there are on average 39 overspeeding incidents per year on the UK mainlines. Hence the statistics from Tyne & Wear Metro is relatively high (75 incidents in two years). However, the high concentration of trainstop equipment contribute to that where the mainlines mostly depend on occasional on train data recorder downloads and speed gun checks.

Localisation of the overspeeding incidents was expected as not all of the stations are equipped with speed control magnets. This is also a reason for location-based correlation with category A SPaDs and trackfault ATP incidents. The worst performing locations for overspeeding incidents all have low approach speed limits, which are between 10 and 20 km/h. However, due to irregularities in speed magnet measurements the drivers have been advised to travel 5 km/h under a speed limit in control locations from the 2012/13 operational year. The Type 1 terminus station with the highest overspeeding incidents stands due to a very steep speed limit drop on the approach (from 80 km/h to effective 5 km/h). Moreover, the distance, which needs to be travelled at 5 km/h, is rather long. During the semi-structured interviews, the drivers complain about frustration, which is caused by some of the incorrect trainstop activations. Furthermore, they revealed that it is hard to keep the required speed after steep drops in speed limits. This is a known phenomenon in automotive industry called ‘speed adaptation’ where drivers tend to underestimate their speed after significant reduction in speed limits (Denton, 1976).

Based on the workshop results, many of the overspeeding incidents were revealed to be memory lapses when drivers forget that 30 km/h speed limit is imposed after resetting a driver’s desk. The analysis of the additional indicators confirms that the overspeeding incidents can be caused by decrease in arousal levels. Such decrease is known to lead to decrease in memory performance (Choi *et al.*, 2013). It is important to note that very high arousal levels, usually associated with stress, also lead to decrease in performance. However, lack of association with two hours into a driving portion for the four worst performing locations means that other factors are more influential there. The correlation analysis revealed associations with other station-related incidents in terms of time. Hence it is possible to assume that common causal factors are present. Passenger flows is the main station characteristic that changes throughout the day.

Necessity to monitor platform-train interface can cause distractions to the speed monitoring task. Distraction is a well-known causal factor from other safety-critical industries. Overspeeding incidents at Manors are good examples of distraction adverse effects. Those only happened during major engineering works in the area. This meant a lot of non-routine operations, e.g. interacting with a train dispatcher. The workshop discussion revealed that many drivers exceeded the speed limit after being distracted by a conversation with the dispatcher.

### *5.4.3 Station overruns and failures to call*

The association found between station overruns and failures to call despite only three cases of the latter incident type highlights potential influence of zero-inflated samples and insufficient number of incident reports for such analysis. However, the failures to call happen at locations where station overruns are observed too. In general, station overrun is a precursor to failure to call incident or category A SPaD, depending on a position and an aspect of a signal. Drivers are instructed to miss a station if they realise they will not be able to stop in time without emergency braking. This allows avoiding passenger injuries on expense of their convenience. The association between two incident types is connected with low rail adhesion season. On the other hand, the LRA conditions are usually present during morning and night hours due to a combination of a dew and foliage on tracks, whereas 50% of the failures to call happen during the day. Day time failures to call are potentially linked to unusual station layouts as shown by Northumberland Park. This is Type 2 station with one of the running signals on a platform side and platform side change compared to previous stations. This is also a location where a failure to call happened outside of the normal LRA time. It is possible that design deviations can distract drivers, who require time to acquire SA for an unusual layout, thus leaving some individuals without enough time to stop safely.

Station overruns follow similar hourly distribution to failures to call. Having a bigger sample of incidents, this category shows correlation with potential decrease in drivers' arousal levels. It is possible to assume that similar correlation would be found for the other incident type if not for a small sample size. Weather conditions causing LRA and dispatch equipment faults are similar (fog, mist), hence associations found with both incident types. More overruns happening in the morning is clearly related to

LRA but also could be a case of not sufficient information on rail condition being provided to train drivers. The semi-structured interviews revealed that the drivers normally know where and when to expect wheelslip but no daily briefs or live information is provided. For the worst performing stations, the incidents are spread throughout a day. Two of these stations deviate from standard Type 1 design by not having a running signal right next to the platform edge but approximately 150-200 metres down the line. Even though the drivers are not required to check a signal on approach, this is one of the mechanisms of building up situational awareness of system state (according to the driver interviews). Hence, it is possible for the drivers to create associations between a signal and a stopping position. If so, the drivers might be using the signals as a reference point with deviations creating serious issues. However, the drivers approached during the semi-structured interviews do not confirm this relationship. Another feature of these two stations is one of the longest and straightest approaches, which are rather monotonous (either wall of vegetation or fields). Even though the correlation with this indicator is not statistically significant, the coefficient is positive compared to most of the other driver-related incidents.

### *5.4.4 Passenger entrapments*

Passenger entrapments were one of the targets of the better incident reporting campaign hence the increase in the numbers of incidents towards the end of 2012/13 but it might not be the only cause. The drivers expressed an opinion that most of the incidents are actually passengers' violations who try to board a train after a door closing warning tone starts. In 2012/13 a proportion of the refurbished metrocars grew significantly in the second half of the operational year. One of the major changes introduced in the new metrocar design is related to a door closing procedure. Old design required a driver to press a button to sound a warning tone and then press it again to start the closing sequence. The refurbished design does not require the second step. This leaves the drivers with less control over a situation for entire duration of the sequence, which is approximately seven seconds. It is possible that it took a few months for drivers to get used to the procedure too.

Often happening towards the second part of the year, these incidents are potentially affected by amount of clothing people are wearing in colder months. Metrocar doors are supposed to cycle but might fail to do so if an object trapped is very thin. Furthermore, the semi-structured interviews highlighted that passengers are more inclined to ignore the warning tones in order to avoid waiting for a train at a cold station. An association with dispatch equipment faults in the period-based data set also suggests environmental factors. However, the discussion above should be considered in context of potential skews introduced by the design and safety policy changes.

The daily peak in passenger entrapments coincides with many shoppers travelling, hence the bags, and school children using Tyne & Wear Metro. The latter are renowned for misbehaving at stations and creating dangerous situations while boarding or alighting. There are also more entrapments late at night compared to early morning, due to alcohol-induced anti-social behaviour by passengers. On the other hand, there is no correlation found between number of schools nearby and this incident type. It is possible that the chosen additional indicator does not provide a good estimate for number of pupils using that station, especially when combined with hospital numbers. However, RSSB-funded research showed that children are not a major risk factor in not detecting an entrapment (CCD Design & Ergonomics Ltd, 2005). Traub and Fraser (2013) did not find evidence that children are harder to identify on CCTV but the study used cameras located on a train side.

Statistics suggests poor usability of the monitors used in Tyne & Wear Metro as such locations have three times more incidents than an average station. Drawbacks for this DOO method identified are the age of the equipment, small screen size, occasional poor focus, less flexibility (changing a viewing angle does not help). However, it is sometimes the only viable dispatch method for Type 3 stations or locations with curved platforms. Another feature of built-over (usually Type 1) and underground stations (Type 3) is a rather confined nature of these locations with limited platform width. Such platform width combined with passenger approaches not visible to the drivers creates situations when a passenger can suddenly emerge from a concourse and reach a train within a couple of seconds. Sum of these factors explains why most of four worst performing locations are Type 3 stations.

As one of the few rail systems in the UK solely relying on driver only driving, Tyne & Wear Metro shows necessity to focus on this equipment to maintain its usability. In the semi-structured interviews the drivers never complained about being the only personnel on a train, as long as station features are fit for purpose. The informal conversations with the drivers held in 2015 showed that many usability improvements had been achieved since 2013 by using heated mirrors, and bigger or better monitors. Apart from weather-related factors, the locations of such incidents correlate with stations where foliage obstructions are more pronounced. Associations are found with door obstructions in all three data sets. It further shows how thin the line is between these two incident types.

### 5.4.5 Wrong side door activations

Statistics from LOROL shows a wrong side door activation every 300000 door openings (Basacik and Gibson, 2015). 25 million door openings in Tyne & Wear Metro is an estimate figure for the two years under investigation, which was provided during the semi-structured interviews. Using this number, it is possible to see that a wrong side door release occurs approximately every 550000 door releases. Hence metro drivers perform significantly better than sub-urban train drivers in London. It is possible that the LOROL drivers have selective door opening procedures as another complication or have better reporting culture. On the other hand, shorter station stops can contribute to better metro drivers' performance due to not enough time to fully disengage from station procedures. This is not supported by the correlation analysis with the additional indicators though.

One station that clearly stands out in regards to wrong side door activations is Haymarket. It has 3.5 times more incidents than the second worst performing station. At this location most of the incidents occurred in the second half of 2011/12. Platform side change is not the main contributing factor in these incidents as the most incidents happened at a platform where no side change occurs. The workshop with the safety managers picked up on an interesting detail. As a Type 3 station, Haymarket has advertisement panels on the left hand side wall when arriving to the station. This is the cab side. The spike in the incidents coincided with installation of a new panel right next to a stopping point which many drivers described as having a too bright backlight. The decrease in wrong side doors activations was observed after the panel was changed to a not so bright one. Moreover, the station, due to recent refurbishment, is the brightest underground station in the system. Potential effects of lighting conditions on drivers' performance need to be assessed. For example, Tyne & Wear drivers might expect a platform of an underground station to be the brightest one and use this to make a decision on a side of doors to open subconsciously. If this relationship is violated, as with Haymarket advertisement platform, then they might be struggling to make such decision. Difference in incident rates at Airport (Type 2 terminus) and St James (Type 1 underground terminus) supports a theory about influence of lighting conditions due to uniformity of this parameter at Airport and not at St James.

Approximately 40% of all stations had had one or more incidents of this type during two years under investigation. When only stations where platform change occurs are considered, this proportion rises to 71%. This identifies such design feature as a causal factor. Tyne & Wear Metro tried to address this issue by supplying the drivers with magnets to proactively anticipate platform side for a next station and cover an irrelevant door opening button with it. However, at terminus stations with two platforms Tyne & Wear drivers cannot know in advance which platform the train control system will assign to them.  Unconventional location of a signal might be a causal factor too.

Even though a period-related data is partially skewed by Haymarket, there is higher incident propagation towards the second part of an operational year. Presence of the LRA, personal factors associated with circadian rhythms and social factors potentially distract drivers or decrease their performance during this time of the year. Importance of arousal levels and concentration is further supported by positive medium strength correlation with “dangerous” time into a driver’s shift.

### *5.4.6 Wrong route incidents*

Setting a correct destination code is a simple task. Losses of concentration, vigilance or situational awareness are causal factors for such mistakes. Hence the correlation found with two hours on shift. The workshop revealed that increased reporting in 2012/13 (mostly by SDCs) causes data to be skewed. However, it is possible to assume that data for that year is closer to reality due to such reporting change. Personal factors similar to wrong side doors come into consideration for wrong route incidents due to skew towards autumn/winter months. Finally, for these two incident types peaks in periods 6 and 13 can be potentially associated with engineering works that are usually timed with school and bank holidays. Such works usually include unconventional terminus stations and arriving to a different platform to the one used in normal operations. Times and locations of wrong route incidents mostly depend on nature of these events.

### *5.4.7 Additional indicators*

The increase in incident levels in locations with tunnel exits/entrances on the approach is possibly attributed to how human vision reacts on such changes in environment. Even though there is a potential for rise in arousal levels in such locations (Yang *et al.*, 2012), human eye also requires up to 10 min to fully adapt (Railway Group Standards, 2003), especially when leaving a tunnel on a bright day. The results do not confirm better safety of such locations in urban rail systems.

Distance between the stations is not correlated with any incident types demonstrating limited effects of monotony. Potentially this indicator does not provide good estimate of decrease of arousal levels due to monotony. The Tyne & Wear Metro physical environment is not as monotonous as mainline railways due to a combination of urban and countryside landscapes. Short station runs and limited tunnel sections might not induce monotony.

Most of the driver-related incidents correlate with potential decrease in arousal levels caused by time spent driving. Many of the incidents discussed above require high standard of route knowledge from metro drivers to mitigate against those. As the assumed decrease in arousal levels can lead to decrease in memory performance (Choi *et al.*, 2013), it subsequently affects route knowledge and drivers’ ability to use it.

The number of trains in the network is an important factor in driver-related incidents propagation. However, as many incident types have three peaks, it can be claimed that this causal factor only works in a combination with other factors thus inducing the adverse effects. For example, there are 7.5% less driver-related incidents per station at Network Rail part of the system compare to the central corridor (Pelaw to South Gosforth) whereas the number of train is lower by 75% at certain times of day. Similar logics can be applied to other additional indicators too. Four worst performing stations do not usually show any correlations with the additional indicators suggesting that at those locations local factors may play more important role.

### 5.5 Conclusions

Even though the risks associated with each incident type are different, the incident composition shows how disproportionate is approach which focuses on category A SPaDs. With almost 50% of the incidents associated with station departure and arrival procedures, these incident types are typically less covered by current body of knowledge even though the extent of PTI-related risks in DOO systems is well-known. The different initiative carried out by the operator led to better reporting and improvements to safety culture. However, those can potentially skew the data too but it has been taken into account during the analysis.

The importance of design-related factors has been highlighted, especially in context of routine timetabled operations and route knowledge. Although the features of metro systems suggest better route knowledge and an increased familiarity of drivers with a system, the results show that this can lead to the unsatisfactory safety performance. One of the recurring topics in the discussion is station design deviation from the typical design templates used in Tyne & Wear Metro. This trend, however, has only been observed for locations where all station infrastructure elements are retained. When those are reduced, number of incidents reduces too implying positive benefits of simplified designs. Among design features potentially increasing incident rates are sharp drops in speed limits, changing platform sides, locations of a running signal, type of DOO equipment and confined nature of stations.

Loss of situation awareness (be it from a disruption or non-routine operations) is another important incident propagation factor. Certain locations have design features which would increase probability of depleted SA occurring, e.g. limited advance warning or presence of sidings. Many of these occasions can be prevented or at least limited with design interventions, and Tyne & Wear Metro has been considering some of those already. Other performance shaping factors identified in the analysis are decreased arousal levels, seasonal environmental factors, personal factors (holidays or financial concerns). Similarly, passenger loading shows an important contribution to incident causality through distraction and making PTI-related tasks more demanding.

Drivers were found to lose their alertness and high performance levels approximately two hours into the driving portion of a shift. This finding, when used for rostering, could improve Tyne & Wear Metro’s safety performance even further. Even though the number of trains passing through a station is not directly related to the number of incidents at that location, it can induce incident propagation in the problematic parts of the network.

Even though the findings were interpreted in the context of Tyne & Wear Metro through the workshop and the semi-structured interviews, those only involved a limited number of front line staff at each time. Hence a bigger scale survey of drivers about the PSFs embedded in the system is required to better understand the phenomena.

# Chapter 6. Questionnaire study

## 6.1 Introduction

With the historic incident analysis uncovering trends in driver-related incidents propagation, it is necessary to expand on those findings by addressing drivers' perception of system design. The semi-structured interviews proved useful in Chapter 5 in creating a more systematic understanding of the phenomena. Tyne & Wear Metro drivers have the most experience of interacting with the physical environment of the system. Thus they are the most knowledgeable people to assess both negative and positive effects of various design features. However, raising a single question only with three drivers creates significant biases thus a questionnaire study with a bigger sample is required.

As it is necessary to understand operational context from drivers' point of view if a multi-dimensionality of the research is to be achieved. With the research philosophy requiring pragmatic approach to a problem, mixing of the methods was performed on the same research level. A mixed methods questionnaire was developed focusing on potential issues uncovered in Chapter 5 but also allowing participants to freely express their opinions in open-ended questions. Vitale *et al.* (2008) claim that the mixed method design becomes increasingly popular because of the complimentary nature of the data it provides. Among the benefits of the mixed method design are better insights and an ability to isolate a variable (Borkan, 2004), elaboration of the analysis to increase the level of details and starting new lines of thinking (Vitale *et al.*, 2008).

Questionnaire surveys are an established method in the railway industry to source attitudes from as big as possible sample of drivers. One of the largest questionnaire studies in railway industry was conducted by Ryan *et al.* (2009b) who collected opinions of 4686 signalers on a range of human factors. The study addresses some of the similar issues as this thesis, including human-machine interface, usability and safety culture. However, it does so based on the existing body of knowledge which is not supported by more in-depth trends analysis for the same sample. The historic incident analysis allows targeting more specific design features. Stevenson *et al.* (2000) used a series of questionnaires to assess ergonomics of train control devices in, at that time, the new Tangara train in Australia. Their method was based on performing an identical questionnaire experiment before and after re-designing of the cab environment and subsequently comparing the results. This approach can be very effective when built around a design change but heavily depends on opportunities.

The questionnaire surveys have been used to study praise that train drivers receive (Horishita *et al.*, 2013), motivations of the train drivers (Fujino *et al.*, 2013) and organisational factors influencing decisions to report or not an incident (Clarke, 1998). Questionnaire studies have also been performed to investigate train drivers' job stress (Chang *et al.*, 2005), physiological reactions to on-the-track accidents (Vatshelle and Moen, 1997). The railway industry is using questionnaires not only for research purposes but also in train drivers selection process. Most of the companies include one or several questionnaire tasks in their assessment of the candidates for drivers' positions. Those mostly focus on psychological evaluation of the candidates. Guidelines for psychological assessment of train drivers produced by Community of European Railway and Infrastructure Companies (2009) claim that such tests can help avoiding employing a person who is not suitable for safety-critical role. Furthermore, questionnaire studies have an important role in change management giving the front line staff an opportunity to provide feedback on performance of an organisation, various work-related issues and changes offered.  For instance, Inoue *et al.* (2015) used questionnaires as a self-assessment tool for train drivers and a source of feedback on new training techniques introduced. Self-administered questionnaires were used to explore symptoms of post-traumatic stress disorder among train drivers (Yum *et al.*, 2006). Impact of sleep quality (Jeon *et al.*, 2014) and chronotypes (De Araújo Fernandes Jr *et al.*, 2013) on train drivers' performance has been studied.

The automotive industry has been using questionnaires extensively to evaluate current designs and requirements of car drivers. For example, the questionnaire study by Herriotts (2005) addresses the issue of the car designs being suited for ageing demographics of a driver. In the wider ergonomics field one of the most popular usability tools is the System Usability Scale (SUS) questionnaire developed by Brooke (1986). Even though the author refers to the SUS questionnaire as "quick and dirty", (Lewis and Sauro, 2009) find it to be of high consistency.

There are many commonalities between the questionnaire studies mentioned above. The most wide-spread method is a self-administered questionnaire combining qualitative and quantitative inquiries. Among the quantitative methods, Likert-type questions are popular for sourcing respondents' attitudes on a scale of "completely agree” to “completely disagree”. However, choosing an optimum number of answers on the scale can be critical to the quality of results. Finstad (2010) shows that the 7-point scale provides the best number of options. Further increase of the scale can lead to a reduced response rate (Nunnally and Bernstein, 1994). Additionally, closed design questions with pre-selected answers are typically added to the questionnaires for categorisation purposes.

The mixed method questionnaires, similarly to pragmatists’ mixed methods approach, aim to expand researchers’ insights in an area of study. Open-ended questions were found to be the most popular qualitative technique. Almost each questionnaire reviewed has at least one open-ended question in the end asking respondents to provide any other comments, concerns or ideas on the research problem. The questionnaire surveys reviewed often had a control group in order to source baseline results. For example, the study on accident propagation by Cui *et al.* (2007) targeted respondents with previous accident history but also involved some drivers without any accident history as a control group.

The statistical analysis of the results usually explores relationships between categories (based on categorical questions) and compares the sub-samples on total scores. Most often Pearson’s chi-square and Student’s t-test respectively are used for such statistical analysis. Use of non-parametric analogues of the t-test is quite rare and no information is provided on distribution of variables. In terms of the descriptive statistics, mean and standard deviation are always provided. However, Jamieson (2004) reminds that the Likert scale results are ordinal values, essentially ranks, and thus cannot be treated as interval values despite it being a common practice (Blaikie, 2003). It is incorrect to assume that the interval between the points on, for example, the 5-point Likert scale is always the same. To put it in perspective, a respondent not necessarily feels the same between Strong disagreement and disagreement, and disagreement and no opinion. Hence median and mode values should be reported for questionnaire results (Jamieson, 2004), but statistical analysis should be carried out using non-parametric tests (Kostoulas, 2013).

## 6.2 Methodology

The questionnaire study was conducted among Tyne & Wear Metro drivers in January and February 2015 with an ultimate goal to understand what design features drivers believe to have adverse effect on their safety-related performance. 43 respondents participated in the survey which was almost 30% of the entire Tyne &

Wear Metro drivers population. According to the methodology presented by Desu (2012), sample size for a finite population of 150 should be at least 35 (23%) to achieve 95% confidence level. Design of the questionnaire (Appendix I) is based on the review of common practices in railway industry.

### *6.2.1 Type*

A roster-based nature of metro drivers' jobs require significant amount of time to organise in-person or telephone interviews. Hence it was decided to conduct a self-administered questionnaire study which is distributed in person. This method along with numerous design iterations allows negotiating potential disadvantages of typical self-administered questionnaires, e.g. higher requirements for questionnaire design and lack of immediate quality control (Fowler, 2009). Furthermore, high level of personal relevance and lack of interaction with an interviewer can induce openness and honesty, especially in sensitive questions (Gillham, 2000).

### *6.2.2 Structure*

The questionnaire consists of 15 questions. Mixed methods approach is used in design (Borkan, 2004; Vitale *et al.*, 2008; Creswell, 2014). Firstly, the metro drivers have to assess 27 statements in question 1 on 7-point Likert (Finstad, 2010) scale which address potential PSFs discussed in Chapter 5. Presence of “do not know” answer in the 7-point Likert questions should discourage people from skipping a question. Next, the respondents are asked about usability of different types of signals used in Tyne & Wear Metro. The participants also have to mark potential contributory factors and name the riskiest locations for wrong side door activations and passenger entrapment incidents. In the second part of the questionnaire, they have to assess various elements of the cab design, and own arousal and boredom levels at different intervals into a driving shift. Finally, the respondents have to provide information on previous incident involvement and demographical data. The questionnaire, as it was distributed to the participants, can be seen in appendix I.

The optimum length of the questionnaire is something that much of the literature seems to agree on. Four to eight pages is a preferred length of the questionnaire with 12 pages possible in exceptional cases (Gillham, 2000; Blair *et al.*, 2013). However, the time required to complete it also has to be taken into the account with a maximum duration being 20 minutes. The questionnaire developed is six pages long including the 1-page long introduction. The final version trials conducted before distributing the questionnaire showed that the maximum completion time was 12 minutes.

15 version of the questionnaire have been developed prior to selection of the final design. Numerous trials were conducted with railway researchers and Tyne & Wear Metro personnel to address questions quality and terminology concerns. The final version of the questionnaire which had been used in the study can be seen in Appendix I.

The statements in question 1 (see Table 13) are mixed to avoid the respondents seeing a grouping pattern. The order of questions follows an advice by Blair *et al.* (2013) with the easiest closed format attitude questions, which are applicable to the entire population, in the beginning. The questions had been simplified during the design iterations as complex behavioural enquiries are non-advisable (Gillham, 2000). The demographics questions are located in the end of the survey to avoid the respondents being intimidated by those before they start. Finally, the questions are kept within 25 words limit and double negatives are avoided.

### 6.2.3 Sample

In total 43 metro drivers participated in the research. Out of 43 respondents, 39 are male, two are female and two decided not to state their gender. 42% of the participants fall into 26-35 years old category with the second biggest group being 46-55 (26%). The biggest proportion of the respondents (31%) had been metro drivers for less than three years at the moment of the study, highlighting generation change in the company. These drivers are considered “inexperienced” by Tyne & Wear Metro. Only 40% of the participants had not been involved in any driver-related incident in the three years before the survey.

### 6.2.4 Data analysis

Ordinal data (Likert-scale questions) was assessed using mode and median values. Mean and standard deviation statistics are produced for interval data (marking questions). Wilcoxon Signed Rank Test was performed on the data to check whether differences in mean and median values are statistically significant. The demographics questions allowed categorizing drivers and exploring whether involvement in different categories affects drivers’ attitudes. To do this Mann-Whitney U-test and Kruskall-Wallis H-tests were performed.

Fisher’s exact test (Fisher, 1935) has been selected to test how likely that an observed distribution between sets of data is due to a chance. This is normally done by Pearson’s chi square ($\chi^2$-test) test but the sample size is not big enough for it (Yates *et al.*, 1999). It is not the case for Fisher’s exact test, which however assumes fixed totals (Howard, 1998). As it was impossible to know totals for each category in advance, the results of this test should be treated as more conservative than normal.

Internal reliability of Likert-scale questions is calculated using Cronbach’s alpha (Cronbach, 1951) to check whether it is consistently reflecting a conduct that it is measuring (Field, 2013). Literature shows that acceptable value for Cronbach’s alpha test is 0.7 and more (Cortina, 1993; Field, 2013). All of the statements in question 1 demonstrate high level of internal consistency with the lowest value of Cronbach’s alpha found to be 0.738.  The statements in question 8 show even higher internal consistency (0.836 minimum).

## 6.3 Results

### *6.3.1 Descriptive statistics*

Table 13 below provides descriptive statistics for the 27 attitude statements in question 1. Drivers tend to assess their performance positively whilst not reporting many effects of the design features.

| Statement | Mode | Median |
|---|---|---|
| 1. The safety record of the Tyne & Wear Metro in the past 3 years has improved | Not sure | Not sure |
| 2. Since I joined the Metro, my safety-related performance has changed for the better | Agree | Agree |
| 3. My route knowledge of the Metro is good | Strongly agree | Strongly agree |
| 4. My confidence reduces while driving during possessions or engineering works | Agree | Just agree |
| 5. The training provided for operations in degraded mode is adequate | Just agree | Just agree |
| 6. The moment I enter or leave a tunnel, I feel more alert | Disagree | Not sure |
| 7.Running signals between the stations are easy to interact with | Agree | Agree |
| 8. Running signals at the stations are easy to interact with | Agree | Agree |
| 9. I am less alert if the outside physical environment is monotonous | Just disagree | Not sure |
| 10. I prefer varied outside environment, such as a mix of vegetation and buildings | Agree | Just agree |
| 11. The recent change in door closing procedure from 2 to 1 button sequence is easier to operate | Agree | Just agree |
| 12. A 1-button sequence might increase the occurrence of passenger entrapment | Strongly agree | Agree |
| 13. I like mirrors as station dispatch equipment | Agree | Just agree |
| 14. I like monitors as station dispatch equipment | Agree | Agree |
| 15. When coming to a scheduled stop I pay attention to a running signal at the platform end | Agree | Agree |
| 16. Running signals located far from the platform end can make selection of a stopping position difficult | Disagree | Disagree |
| 17. It is difficult to choose which side doors to open when station signals and the platform are on opposite sides | Disagree | Just disagree |
| 18. Signalling at ground level can be confusing after driving a train in passenger service | Just disagree | Just disagree |
| 19. The change of platform side does not affect my ability to select correct side to open the doors | Agree | Just agree |
| 20. The stations differ a lot in terms of driver visibility of passengers on a platform. | Strongly agree | Agree |
| 21. I prefer the ¾ life refurbished cab to the original one | Agree | Not sure |

| Statement | Mode | Median |
|---|---|---|
| 22. If the time between two stations is more than 2.5 minutes, this improves my alertness | Not sure | Not sure |
| 23. I feel more alert when the time between stations is less than 1 minute | Not sure | Not sure |
| 24. I prefer steep change in speed limits rather than gradual change | Not sure | Not sure |
| 25. My familiarity with operational protocols at sidings is very good | Agree | Agree |
| 26. I have good familiarity with the layout of the depot | Agree | Agree |
| 27. I find it harder to keep within higher speed limit than lower speed limit | Disagree | Disagree |

**Table 13. Drivers' assessment of attitude statements**

Average marks and standard deviation statistics for different types of signalling are summarised in Table 14. The running signals on Network Rail infrastructure are easier to interact with than the Tyne & Wear Metro signals (*Z=-4.464, p=0.00*) and the ground position lights (*Z=-5.005, p=0.00*). The ground position lights got the lowest marks out of all the signalling types in question 2. The mean mark for those is significantly lower than for the Tyne & Wear Metro signals too (*Z=-2.922, p=0.003).*

| Signal type | Mark/10 | St. deviation |
|---|---|---|
| **Running signal on Network Rail infrastructure** | 9.37 | 1.155 |
| **Running signal on Tyne & Wear Metro infrastructure** | 8.19 | 1.694 |
| **Repeater** | 8.44 | 1.623 |
| **Advance warning signal** | 7.88 | 2.184 |
| **Flashing aspects** | 8.98 | 1.520 |
| **Ground position lights** | 7.28 | 1.944 |
| **Junction indicators** | 8.40 | 1.591 |

**Table 14. Marks for different signalling types used in the Tyne & Wear Metro (out of 10)**

When asked to expand on the reasons for allocation of the low marks for the signalling types the respondents mostly mentioned not enough height, poor positioning and overgrowth of the ground position lights. Advance warning signals, which are effectively notice boards, were critiqued on not being distinguishable as much as signals, especially at night. The Tyne & Wear Metro running signals got some critique for being located on bends, in tunnels and having less advance warning than mainline signals.

Table 15 provides similar statistics for drivers' views on potential causal factors in wrong side door activation incidents. The respondents believe that the biggest contributing factor to wrong side doors activations is "attention lapse" with "various distractions” being in a close second place. Furthermore, the Wilcoxon Signed-Ranks test (*Z=-1.650, p=0.099*) did not find the marks for these two potential causal factors significantly different. The respondents feel that “inadequate training” does not have much to do with wrong side doors activation propagation. “Lack of reminders for drivers” also is deemed to be less important potential causal factor. However, the difference between marks for “inadequate training” and “lack of reminders” factors is significant (*Z=-2.684, p=0.007*). The participants did not make much use of the open space to add their potential causal factors. “Difference in button layout” and “changes of platform side” had been mentioned twice each.

| **Potential causal factor** | **Mark/10** | **St. deviation** |
|---|---|---|
| **Attention lapse** | 8.67 | 2.078 |
| **Lack of reminders for drivers at stations** | 4.77 | 2.951 |
| **Layout of the door control** | 6.79 | 2.833 |
| **Distractions** | 8.28 | 1.956 |
| **Inadequate training** | 3.37 | 3.079 |

**Table 15. Marks for PSFs in wrong side door activation incidents (out of 10)**

In the assessment of potential causal factors, the participants were also asked to name three stations that, in their opinion, carry most risk of each incident type and reasons for this choice (Table 16). In terms of wrong side doors activations, none of the four worst performing stations (Section 5.3.2) is named by more than the third of the sample. However, many participants wrote only one or two station names. West Jesmond and St James are not mentioned at all. “Change of the platform side” is mentioned the most as a reason to consider a station risky for this incident type. In terms of “distractions”, only the Living Wall art installation at Sunderland station is identified. Furthermore, Northumberland Park is the only station that is specifically commented on its design.

| **Station** | **Reason** | **Times mentioned** |
|---|---|---|
| **N.Park** | Change of platform side | 7 |
| **Haymarket** | Opposite side on inline | 6 |
| **Jesmond** | Platform 2 – change of platform side | 5 |
| **N.Park** | Island platform | 4 |
| **Heworth** | Platform side change from both directions | 4 |
| **Sunderland** | Platform and monitors are on opposite sides | 4 |
| **Sunderland** | Platform 3 (living wall) | 3 |
| **Sunderland** | First change in side for a long time | 3 |
| **Pelaw** | Change of platform side | 3 |
| **All stations** | Same chances everywhere | 3 |
| **Underground stations** | DOO opposite side to platform | 2 |
| **Gateshead** | No reason stated | 2 |
| **Sunderland** | “Walking passenger” wall – distraction | 2 |
| **Felling** | Platform 2 (personally) | 1 |
| **Manors** | Platform 1 (inexplicable, could be by chance) | 1 |

**Table 16. Drivers' analysis of the worst performing locations with contributing factors for wrong side doors incidents**

Similarly to wrong side door activations, the respondents had assessed a list of potential causal factors for passenger entrapment incidents with the descriptive statistics summarized in Table 17.

| **Potential causal factor** | **Mark/10** | **St. deviation** |
|---|---|---|
| **Night time** | 6.05 | 3.124 |
| **Snow** | 3.63 | 2.664 |
| **Rain** | 4.60 | 2.727 |
| **Mist** | 6.35 | 2.811 |
| **Direct sunlight** | 8.21 | 2.253 |
| **Vegetation overgrowth** | 4.65 | 3.085 |
| **Location of station infrastructure, e.g. CCTV cameras** | 8.09 | 2.147 |
| **Overcrowding at a platform or in a train** | 9 | 1.512 |
| **Winter clothing on passengers** | 4.60 | 2.896 |
| **Shopping bags and suitcases** | 6.37 | 2.691 |
| **Layout of the stations** | 6.58 | 2.684 |
| **Design of passenger approaches** | 7.49 | 2.576 |
| **Mobility aid equipment, e.g. walking sticks, crutches** | 7.33 | 2.378 |
| **Station dispatch instructions/procedures used in the Metro** | 5.12 | 5.72 |
| **Low height passengers, e.g. children** | 2.72 | 2.684 |

**Table 17. Marks for PSFs in passenger entrapment incidents (out of 10)**

The highest marks are given to “overcrowding at a platform or a train”, “direct sunlight” and “location of station infrastructure” in this particular order. The Wilcoxon Signed-Ranks test returns significant difference between *the “overcrowding” factor* and the other two highly marked factors (*Z=-2.682, p=0.07 and -3.925, p=0.00* *respectively*) however no significant difference between the “direct sunlight” and “location of station infrastructure*”* factors (*z=-0.426, p=0.67*). “Design of passenger approaches” and “use of mobility equipment by passengers” are scored relatively high too. The participating drivers do not see “snow”, “rain”, “winter clothing” and “vegetation” as major contributing factors to propagation of the passenger entrapment incidents. When asked to add other contributing factors, 26% (11 out of 43) of the metro drivers include “late boarding behaviour from passengers”. Moreover, the related behaviour when people are “holding doors for late passengers” is mentioned in 14% of the questionnaires.

For passenger entrapments incidents, the respondents are better aware of the historic incident data with one of the worst performing stations (Haymarket) mentioned in more than 50% of the questionnaires. On the other hand, the only type 1 station is not mentioned at all despite having high passenger entrapment incident rates. Main issues identified by the drivers are “the age and quality of the DOO equipment” at Type 3 stations, using monitors as dispatch equipment. Furthermore, “overcrowding” and “lack of passenger visibility on approaches” are named as main risk factors for the worst performing stations. Finally, poor lighting at a back of a platform and curved platforms were named as potential passenger entrapment causes. The former PSF is mentioned in relation to Byker (Type 1 built-over) station as the riskiest location even though no passenger entrapments were registered at this location in 2011/12 and 2012/13 operational years (Appendix C). The participants mostly do not know what is the season with the highest number of passenger entrapments. Of those who believe to know, Summer is the most popular answer.

The participants are mostly just satisfied with the various elements of the cab environment (Table 18). The negative attitudes are observed towards driver's desk and, especially, HVAC unit.

| Component | Median | Mode |
|---|---|---|
| **Driver's seat** | Just satisfied | Satisfied |
| **Driver's desk** | Just dissatisfied | Just satisfied |
| **FASSI** | Just satisfied | Satisfied |
| **Driver's Safety Device (DSD)** | Just satisfied | Satisfied |
| **Deadman's Vigilance Device (DVD)** | Satisfied | Satisfied |
| **Master controller** | Just satisfied | Satisfied |
| **Heating, ventilation and air conditioning (HVAC) unit** | Dissatisfied | Strongly dissatisfied |

**Table 18. Drivers' assessment of the elements of the cab (7-point Likert)**

The statistics for the drivers' assessment of their boredom and alertness, and overall incident occurrence risk can be seen in Table 19. The questionnaire study demonstrates that the majority of the metro drivers believe that they start to get bored and become less alert at least after two hours of constant driving. They also think that the beginning of the shift has less risks for incident occurrence but driving time of more than 3 hours increases incident rates.

| | From the start of the fixed driving portion, in minutes | |
|---|---|---|
| **Statement** | **Mode** | **Median** |
| **My alertness starts to decrease** | Over 181 | 151-180 |
| **My boredom starts** | Over 181 | 121-150 |
| **Most incidents are likely to happen** | Over 181 | Over 181 |
| **Least incidents are likely to happen** | <30 | 30-60 |

**Table 19. Drivers' assessment of their boredom and alertness levels**

### *6.3.2 Influence of the respondent's age and experience on their attitudes*

The only significant age difference is identified in attitudes towards the design of the refurbished cab (H = 9.805, p = 0.020). The drivers who fall into 46-55 years old age band mostly dislike the new cab whereas the 36-45 and 55+ age groups mostly prefer it. In terms of experience, the only significant relationship (at 95% confidence level) is related to choosing of a stopping position. More experienced drivers tend to disagree less that it is harder to select a stopping position when a running signal is far from a platform's edge (H = 8.903, p = 0.031).

## 6.4 Discussion

### *6.4.1 Safety-related performance*

In general, and based on the results described in the previous section, it can be concluded that Tyne & Wear Metro drivers have no concerns of any significant effects on their safety-related performance produced by various features of infrastructure design. The drivers believe that their safety-related performance is improving whereas route knowledge is already very good. This is in line with previous research reporting great importance of experience and route knowledge in train driving process (Naweed, 2013). As Tyne & Wear Metro can be considered a system with a low automation level under the model proposed by Parasuraman *et al.* (2000) it is suggested that disturbances in the system can lead to dips in performance (Balfe *et al.*, 2015; Madigan *et al.*, 2016). The results support the assumptions of significant dips in safety-related performance in non-routine operations (Section 5.5). Drivers do not feel as confident in such events, which could be caused by an over-reliance on route knowledge. The habituation and expectation bias are developed in repetitive experiences and cause loss of the SA (O'Connell *et al.*, 2015).

The respondents indicate there is a room for improvement in terms of training for degraded mode. As of 2015, Tyne & Wear Metro was using a SimKit method for simulation of different scenarios. The SimKit itself resembles a board game with a map of the system and additional figurines (trains, people, signs and signals) for different operational situations. The drivers are presented with a scenario on the map and asked to explain their actions in such situation. Looking back on the questionnaire results, it could be a case of the SimKit used just not providing enough association with a simulated event for Tyne & Wear Metro drivers.

Even though the past incident statistics demonstrates that safety culture is improving in the Tyne & Wear Metro, drivers do not feel improvements in the safety record. It is possible that they do not associate statement 1 and the safety culture of the organization but the median scores are similar to the safety culture assessment scores of other railway staff in the UK (Ryan *et al.*, 2009a). The results corroborate findings of the historic incident analysis about lack of relationship between tunnels and arousal levels.

### *6.4.2 Signals and the related factors*

Presence of the SASSPaDs in the historic data suggests a complexity of interaction with signals from a stopped train. Such interaction usually occurs at a station or in sidings. However, the drivers do not believe the station signals being more complex than the signals between the stations (Z = -1.155, p = 0.248). On the other hand, the respondents confirm that the GPL, a signalling type mostly present in the sidings, is the least usable by far even though question 1 shows no confusion with such signals and the associated areas. It is possible that the respondents cannot associate themselves with statement 18 due to the second part of it ("…*after driving a train in passenger service*") but can discriminate between the signalling types taken out of context. Similarly, the fact that the drivers see the difference between the signaling types might not mean that the worst scored type is uncomfortable to use. Nevertheless, Multer *et al.* (2015) identified problems with the GPL sighting from certain cab. Moreover, this signal type definitely fails comfortable signal height parameters outlined by Li (2003).

The signals at Network Rail part of the network have been assessed to have better usability than the Tyne & Wear Metro signals. This supports previously proposed hypotheses on such signals providing better advance warning and subsequent situational awareness. Moreover, less curved geometry and larger distances between the signals potentially add some marks to the mainline 4-aspect signals too.

### *6.4.3 Passenger entrapment factors*

Higher localisation of passenger entrapments to underground stations shows problems with predominant dispatch equipment at such stations – monitors. Moreover, the drivers' list mostly underground stations as the most entrapment prone. Yet the median for mirrors' assessment is less positive than the median for monitors' assessment. The Wilcoxon Signed-Ranks test supports statistically significant difference (Z = -3.242, p = 0.001) thus not confirming poorer usability of the DOO monitors. The Tyne & Wear Metro operator had dedicated a lot of resources to improvements of DOO equipment between 2013 and 2015. Moreover, most of the resources were allocated to locations with monitors as DOO equipment. This could have caused the discrepancy between the historic data and the questionnaire results. Basacik *et al.* (2009) only list advantages of CCTV in the assessment of DOO disregarding mirrors. However, it is possible that their research project means individual door cameras with the term CCTV. On the other hand, CCTV-related DOO is only one of the design features potentially acting as adverse PSF at underground stations.

Table 17 demonstrates that passenger levels (*overcrowding*) associated with the underground stations as well as layout of such stations could be the factors increasing amount of passenger entrapments. Moreover, the semi-structured interviews with drivers showed that at some locations, e.g. Haymarket, monitors are not fit for purpose only in combination with crowded platforms as a view of some doors is obscured when a platform is full. Poorer usability of the mirrors is reinforced by the mark for *direct sunlight* as most of the overground stations have the mirrors as DOO equipment. This also makes *direct sunlight* the only important environmental factor as the rest were marked relatively low. It corroborates the Tyne & Wear Metro operator's risk assessment of some stations with the mirrors where dispatchers had to be introduced for the hours of direct sunlight. Drivers commented a lot on passenger behaviour as a contributory factor. According to Basacik *et al.* (2009) lack of staff presence in DOO systems encourages unsafe acts.

Confined nature of the underground stations does not allow good visibility of passengers on approach. The car driver study byYoung *et al.* (2015) demonstrated that built environment is not as an important mental workload inducing factor as high pedestrian/traffic density. On the other hand, the scenario in that experiment did not require drivers to extract safety-critical information from the built environment. The recent change in the door closing procedure leaves Tyne & Wear Metro drivers with less control over passenger entrapment risks at locations with poor passenger approach visibility. This is supported by the assessment of statements 11 and 20.

#### *6.4.4 Contributors to station overruns*

Statements 15 and 16 in question 1 represented station overrun incidents. RSSB research (Roels and Mills, 2010) claim that errors and performance shaping factor can be encountered in 3:1 ratio in station overrun incidents. Environmental factors are responsible for approximately 15% of station overruns/failures to call in that research. However, that study does not consider any design-related PSFs. Moreover, mainline trains are often timetabled to miss certain stations hence importance of remembering a correct diagram and personal factors is higher. Based on the station overrun statistics (Section 5.4.3) it is possible that a running signal's position could affect the drivers' train stopping performance. Even though it is not a step in the mandatory stopping routine, the participants admitted checking a running signal before a station stop. It is possible that these gazes on a running signal are purely a situational awareness exercise. Checking a running signal creates an additional pressure on already limited visual attention, which might hinder drivers’ ability to select a correct stopping position on time. However, it is unknown whether ease of interaction with other elements of station design is affected by the location of a running signal. The drivers do not believe that the position of the signal has any effect on their stopping performance.

### *6.4.5 Overspeeding-related factors*

Historic overspeeding statistics from Tyne & Wear Metro shows that the stations with the lowest speed limits are prone to propagation of overspeeding incidents (Section 5.4.2). This is corroborated by drivers’ attitudes towards different speed limits. The respondents have stated that they find it harder to keep to a lower speed limit than a higher one. This potentially corroborates with the lower mark for the GPL, as associated areas usually have low speed limits. This assessment also corroborates added complexity of procedures employed by Tyne & Wear Metro, especially the “5 km/h under” initiative. It is important to note that 5 km/h is 50% of the allowed speed limit at some of the stations whereas at different locations it could be only 10-15% of a speed limit. Moreover, the low speed is perceived even lower after driving at a higher speed (Denton, 1976). There has been a spread of opinions when the participants were asked to express their attitudes towards gradual versus steep change in speed limits, potentially showing that the rate of decrease might not be an important PSF.

### *6.4.6 Wrong-side doors activations and related factors*

Statements 15, 17 and 19 are addressing wrong side doors activation incidents. The respondents disagree that the location of a running signal affects their door opening performance. They do not think that the change of platform side does it either. This contradicts the previous findings by Basacik and Gibson (2015) and Chapter 5. On the other hand, 67% of the stations named to be associated with the wrong side doors activations are the stations where a platform side change occurs. It is possible that the drivers, similarly to the situation with the GPL, can discriminate between different stations but still do not see these locations as risky. Furthermore, the wording selected for statements might have caused such controversy. In case of statement 17, it is possible that the participants saw it as a question about the underground stations with an island platform (type 3) instead of a question about the stations deviating from a standardized design, e.g. Tynemouth (type 1) and Northumberland Park (type 2).

Previous research indicates a possible cause to wrong side door releases the transition of door opening skills from rule based to routine based domain over the course of a driving portion. This leads to a subsequent failure to notice change of a platform side (Halliday, 1995). It is corroborated by drivers' high marks for *attention lapse. Distractions* have been considered as one of the main causal factors too (table 3). The unusual cab design means that having people sitting next to the cab is almost similar to having someone else in a normal sized mainline cab, and this is known to cause distractions (Verstappen, 2015). Immediate (cab) and outside physical environment can present many distractions that are known to interfere with a safe procedure (Caird *et al.*, 2002; Edquist *et al.*, 2011; Johanning, 2011; Salmon *et al.*, 2011; Young *et al.*, 2015). Many distractions are caused by the noise from passengers sitting next to a cab due to the unusual metrocar design. Lack of reminders received low marks showing that the respondents are happy with the quality of stopping position markers and a few of platform side indicators.

### 6.4.7 Drivers' assessment of their arousal and alertness levels

The results do not confirm relationship between tunnel exit/entrance and increase in driver's arousal levels (Yang *et al.*, 2012) thus corroborating previous findings from the historic incident data (Section 5.4.7). No effects of distances between stations have been noticed by the respondents hence supporting previous findings too. Despite found links between monotony and increase in safety risk (Thiffault and Bergeron, 2003; Williamson *et al.*, 2011), the drivers did not have a strong opinion on effects of monotonous outside environment on alertness levels. There is a possibility that the drivers see the entire environment along the routes as monotonous due to repetitiveness of their daily task. On the other hand, they still like the varied environment more thus showing that they can distinguish several types of the outside environment.

The drivers' assessment of their boredom and alertness (Table 19) also supports the previous work (Keun Sang and Ohkubo, 1994; Folkard, 1997; Macdonald *et al.*, 1997; Yang *et al.*, 2012) as the participants claim to start losing alertness and developing boredom after two hours of driving. It is important to note that the mode answer for the first two statements in Table 19 is "over 181". Over the course of the research, the drivers have been rather vocal about their concerns and dissatisfaction with long driving hours hence it could have affected their answers. It is possible that if there was an option such as "over 241 minutes" many would have selected it in order to make a point about too long driving portions. The statistics reports that there is no difference between the assessment of the most incident-risky time and the assessment of the time when the alertness starts to decrease (Z = -0.070, p = 0.944). Similarly, no difference has been found with the time the boredom starts (Z = -1.353, p = 0.176). Hence, it is possible to state that the metro drivers link the alertness and the boredom with the incident propagation.

### 6.4.8 Cab design

The questionnaire results do not clearly indicate whether the refurbished cab is accepted more than the original design. Significant differences in attitudes towards the cab from various age groups are perplexing as there is no easy explanation apart from difference in methods and approaches used to train these various groups of drivers. Even though the previous research demonstrates correlation between use of DAS and higher workload (Large *et al.*, 2014) the drivers expressed positive views on FASSI – the DAS used in Tyne & Wear Metro. The non imposing nature of the DAS used in Tyne & Wear Metro potentially contributes to such a positive assessment.

The main areas of dissatisfaction with the new cab are the driver's desk and the HVAC unit. Ambient temperature in a cab can be an important causal factor for drops in performance and vigilance (Human Engineering Ltd, 2005; RSSB, 2008; Ružić and Časnji, 2011). The HVAC unit has been taken off the most trainsets at the time of the survey to be changed for a simple ventilation system. It happened due to unit's inability to maintain a stable comfortable temperature. Consultations with the drivers and safety managers have revealed that the dissatisfaction with the desk can be attributed to differences in design of this element of a cab across the train fleet. For example, location of door controls can differ from cab to cab, which is mentioned several times in the questionnaires in regards to wrong side door activations. Moreover, it is not very ergonomically sound as some procedures, e.g. decoupling, do not take into account a potential variation of physical parameters across the metro train drivers' population. Basacik *et al.* (2009) in their assessment of the DOO dispatch method emphasise that door controls compete for drivers' visual attention. This is further pronounced in systems where the position of these controls is irregular across the fleet.

### 6.4.9 Awareness of the incident statistics

Even though the assessment of the statements and the answers demonstrate positivism about training provided to the drivers, poor awareness of the incident statistics transpires through the questionnaires. For example, the lack of knowledge about the worst performing stations for passenger entrapments and wrong side door activations. Hence, the drivers cannot gain additional safety benefits based on experience and knowledge at those locations. On the other hand, excessive focus only at certain areas can draw drivers' attention from other stations thus creating additional risks there. This can also result in reverse response as shown in Section 5.4. Finding a right balance between incident statistics awareness and general safety training is very important for improvement of the metro safety.

It is possible to claim that personal associations and previous experiences affect the assessments of the risk of certain stations. Passenger entrapment incidents at Byker and Monkseaton (both Type 1) are very good example of this situation. Even though both stations had roughly the same amount of incidents (nine and ten respectively) in a two-year period (2013-2014) before the questionnaire study, the first station is mentioned by 16% of the participants whilst no one comments on the second station. This can be explained by the fact that there has been a high profile incident at Byker where a driver was prosecuted afterwards. The incidents at Monkseaton did not have similar consequences.

The drivers also struggle to name correctly trends in incident propagation. It is possible that trends had changed between 2013 and 2015 but most of the respondents do not even attempt to answer this question. Even though the driver coaches claim to provide this information to drivers during training, there are doubts about consistency of such information due to lack of a structured approach.

### *6.4.10 Effects of individual factors*

Age does not have effect to the same extent as only one statistically significant relationship was found. Even though train driving is a highly visual task and more experienced drivers are known to have more advanced gaze-scanning patterns (Underwood, 2007; Young *et al.*, 2015), only one statistically significant relationship was found between experience and attitudes to the statements. This is different to findings by Ryan *et al.* (2009a) who showed that railway controllers' experience has significantly more effect on their perception of their job.

### *6.4.11 Assessment of the questionnaire results*

The questionnaire results demonstrate that Tyne & Wear Metro drivers do not have concerns about effects of design of the physical environment, considering the current operational procedures as sufficiently safe. Nonetheless, the literature advices that the next step in state-of-the-art accident causation research will be an in-depth look in what is considered “normal” operations and performance (Salmon *et al.*, 2015). Something considered safe can be deemed risky under a scrutiny. Furthermore, as with most questionnaire studies, there are always concerns about respondents’ honesty and openness even though a lot has been done to induce it in this survey.

Even though the results mostly corroborate the findings of the past incident analysis, several unusual results were found. Those can be mostly attributed to the participants’ assessments of the statements and their answers in the rest of the questionnaire. It is possible that the statements were not worded correctly. The attempt to vary wording of similar questions might have introduced a difference in ways how the participants perceived a situation described in a statement. Furthermore, it is possible that the drivers have failed to associate themselves with some statements due to lack of particular experience, survey time constraints and other factors. The above expressed concerns about openness also apply here as the participants could have been biased in order to show their confidence and good level of knowledge. When the questions were non-personal, e.g. marking questions, the drivers might be more inclined to give unbiased answers.

Fisher’s exact test was used to see whether answers in the marking questions influence participants’ attitudes towards some of the statements. Even though several statistically significant relationships exist none of the expected ones are significant, e.g. relationship between the GPL marks and statements 18, 25 and 26. This demonstrates that the drivers approach these questions with separate mind-sets and see different underlying structure to those.

It is important to explore ways how to investigate drivers’ interaction with the physical environment without biases introduced by drivers’ willingness to express certain agenda or lack of it. When assessing the questionnaire, a driver has enough time to consider what his/her answers can achieve or what consequences those might have. Moreover, a retrospective assessment of a situation can be different to a real-time one. Hence methods that would provide insights into drivers’ performance in real time should be explored. Moreover, these methods should be non-intrusive and facilitate real-time driving process without major alterations in order to avoid different types of statistical errors. One of the methods, which suits this description is eye-tracking when drivers’ performance is assessed based on their gaze patterns.

## 6.5 Conclusions

This chapter presents a collection of drivers' attitudes towards specific design features implemented in Tyne & Wear Metro. Many of these elements are used in other urban rail systems in one form or another which allows transferability of the results and methodology. Some of the elements are considered for implementation, e.g. DOO operations. The results provide good overview of drivers' assessment of such features and associated PSFs.

The approach of providing drivers with specific hypotheses based on previous research helps identifying links between different PSFs. The questionnaire results show high level of agreement with the hypotheses set from the historic incident data. Moreover, the findings are often in line with wider human factors research corroborating similar failure mechanisms to other safety related industries. Correlation can be seen between the drivers' selected causal factors and results of previous work in the area. The participants confirm that many features of the immediate physical environment in Tyne & Wear Metro have relatively low satisfaction levels. The main differences with the previous human factors research are found in areas of monotony of a task, effects of experience levels and tunnel exit/entrance.

Even though the drivers generally do not believe that there is a risk associated with the offered statements, it is definite that the respondents can discriminate between elements of the system design in terms of ergonomics. There are concerns about drivers' being able to answer "personalised" statements with full honesty due to discrepancies in assessment of different elements. However, it is possible that even the lowest marked elements are still considered safe but not in comparison with other elements.

Researchers in human factors start focusing on "normal operations" in order to reveal potentially hidden incident causation factors. Discrepancy between some contextual statements and marks demonstrates that drivers might not be able to assess the situations offered in the questionnaire due to method limitations. This strongly implies that further work should be pursued on a cognitive level, where drivers' statements (from this chapter) are compared to their actions. Such comparison will not only provide an additional dimension to the above results but also can reveal hidden causal mechanisms which could be overlooked in the questionnaire study. Moreover, it is necessary to understand whether drivers' discrimination between similar locations is driven only by personal and organisational factors, or there are small design discrepancies which affect that too.

The results highlight a number of areas with future research potential. In terms of design features and with historic incident composition in mind, station layouts show high potential of safety improvements if associated PSFs are investigated. Different station types, positions of running signals and DOO equipment are some of the areas, which should be prioritised for research on a cognitive level. Other avenues to explore include environmental effects on DOO infrastructure, reduction of noise and other distractions in a cab, door controls. Sequence and duration of speed limits should be given consideration too. Signalling types need to be studied further in terms of usability and ease of use. Passenger levels and associate distraction show significant importance on drivers' performance in different (not only station-related) safety-critical tasks.

The questionnaire demonstrates importance of drivers' personal experiences in assessing different situations. Previous incident involvement often makes drivers more critical towards adverse effects posed by the physical environment. It is important to learn more about this relationship in order to explore its potential inclusion into a driver training process. Furthermore, this work shows that metro systems could benefit from raising awareness of the risk-bearing locations among the front line staff. Such awareness should improve the existing route knowledge but should not be overbearing to avoid shifting a focus of the safety-related performance only to certain poorly performing locations. Finally, concerns about drivers' experience and confidence with a degraded system and non-routine operations are corroborated. It is clear that route knowledge can be reinforced by driving experience in a degraded system as well as improved situational awareness.

# Chapter 7. Eye-tracking investigation

## 7.1 Introduction

The results of Chapters 5 and 6 have identified potentially hazardous design features in Tyne & Wear Metro. However, the discrepancies between the historic incident statistics and drivers' perceptions require assessing some of those results with more objective measures. The naturalistic design of the thesis implies the assessment to be carried out in the environment as close to the real operational environment as possible. Hence, non-intrusive methods providing objective data on drivers' interaction with physical environment, preferably on a cognitive level, have been considered.

In selection of a correct methodology, it is necessary to remember that train driving is a highly visual task (Naweed and Balakrishnan, 2013; Naumann *et al.*, 2016). A train driver has to scan approaching physical environment for presence of hazards, warnings and imposed limits. This visual interaction between the driver and infrastructure means that the driver should not only see things but also be able to interpret and process those. Hence visual attention with its several domains is very important (Sturm *et al.*, 1997). The domains are alertness, vigilance, selective and divided attention. Vigilance depends on a person's ability to sustain alertness for prolonged periods of time to detect infrequent but relevant stimuli. However, long lasting tasks with frequent stimuli require sustained attention. Some researchers treat vigilance and sustained attention as the same domain (Strauss *et al.*, 2006; Cheng *et al.*, 2011). Sturm *et al.* (1997) define selective attention as the ability to focus only on certain features of a task while ignoring the rest of the task. Divided attention is an ability to spread the focus capacity across several tasks simultaneously.

Task associated with train driving require the use of all of the attentional domains, i.e. monitoring the track for warning stimuli (signals, speed limits), being able to prioritise or divide focus based on the route knowledge, a rule book and personal experience. However, as indicated by the drivers' responses, some design features might be harder to process hence fixation duration on those is expected to increase. Cheng *et al.* (2011) mention that hazardous situations arise from distractions of the visual attention. Such distractions can be internal or external based on variety of factors including concurrent tasks, workload, personal factors, experience and physical environment. According to Castro (2009), 90% of road traffic accidents happen due to issues with extraction of visual information and, in particular „*I looked but did not* *see it*” scenario. Castro (2009) also concludes that information on visual behaviour of drivers and subsequent analysis provide a powerful tool to quantify attentional processes.

### *7.1.1 Eye-tracking*

Jang *et al.* (2014) claim that the data on eye movement can be a window into human cognitive processes as such movements are controlled by cognition. This statement is based on original eye-mind theories such of Yarbus (1967), and Just and Carpenter (1976); (1980). According to these influential studies, person’s eye fixations on an object mean processing of that object by person’s cognition. As humans interact with visual world they are limited with capacity of human cognition and perception. Hence selection of most relevant information is happening based on current demands of a task. This selection is primarily taking place via the eye movements and is driven by top-down factors (Henderson *et al.*, 2013; Borji and Itti, 2014). Patterns of eye-movements are task-specific and “*allow diagnostics of a task an observer is trying to perform*” (p.788) (Borji *et al.*, 2015). Complex design features increase visual demand and subsequent workload thus distracting from other tasks and altering visual patterns. The altered visual patterns can lead to train drivers neglecting safety-critical elements. Moreover, higher strain on visual attention can lead to the drivers being left with less time to extract information from other elements.

The connection between human gazes and cognition, along with technological advancements, created a new field of human factors research that uses eye-tracking techniques as a non-intrusive way to study human performance and habits. An important advantage of the eye-tracking research methodologies is that fixations are involuntary and are not affected by instructions given to participants making the trials as close to reality as possible (Martens, 2000). This means opportunities for non-biased data collection which is not affected by company safety culture, initiatives. Portable eye-trackers provide realistic and unencumbered interaction with the environment even in hospital emergency rooms (Szulewski and Howes, 2014). Eye-tracking technique has been used in various safety-critical industries to study human behaviour, e.g. risk assessment and hazard identification by motorcycle drivers (Pradhan *et al.*, 2005; Hosking *et al.*, 2010), influence of experience in lane changing behaviour among car drivers (Underwood *et al.*, 2002), human-computer interaction (Jang *et al.*, 2014), path scanning strategies by aircraft pilots (Ottati *et al.*, 1999).

Eye-tracking is a rather novel technique for the railway industry. Perhaps, a slow start in HF research also caused a lag in novel methodologies intake. Simmons (2015) claims that the rail industry had been conservative towards new technology in the past due to its "safety first" culture and budget constraints. However, the need to increase passenger throughput is changing the industry. Moreover, the technology was rather bulky and seemed intrusive at first, which is an obstacle in rather confined cab environments. It became less of an obstacle only in 21st century with increase in portability of the devices. In comparison, active research in car drivers' gaze behaviour started from 1970s (Kapitaniak *et al.*, 2015). Nevertheless, several eye-tracking studies conducted in rail industry. Groeger *et al.* (2001) and RSSB (2005) studied mainline train drivers' visual strategies and reviewed minimum signal reading time through a series of eye-tracking experiments. However, RSSB data was summarised for 10 different routes meaning different PSFs were involved. Moreover, it only considered in-between stations driving. RSSB (2007b) re-used the previously collected eye-tracking data to study design features of 1 signal which had been passed at danger on numerous occasions. The study highlights poor sighting distance and others elements of physical environment competing for visual attention at that stretch of a line. Naghiyev *et al.* (2014) used eye-tracking to study introduction of in-cab signalling on train drivers' visual strategies but the two experimental groups were compared on the basis of different routes.

The number of eye-tracking studies in railway industry is small. It is nowhere near automotive, education and marketing research domains. This comparison clearly shows that the potential of eye-tracking has not been fully exploited yet in train driving research, particularly in relation to urban rail systems. The existing studies are mostly exploratory with many static (in relation to a driver) in-cab elements which is not representative of DOO driving. Milleville-Pennel *et al.* (2010) handled dynamically moving environment by utilising wide AOIs covering approximately 50% of a visual field and disregarding participants' ability to move their heads. More driver head movements are expected in DOO systems.

All of the above railway eye-tracking experiments happened in conjunction with other experimental methods. For example, Naghiyev *et al.* (2015) evaluate results of their eye-tracking study through a multi-step workshop. De Ceunynck *et al.* (2015) claim that the eye-tracking research has to be supported by other methods. The modern approach to eye behaviour assumes that the correlation between processing and fixations is not perfect, and there might be a lag between those (Deubel, 2008; Beanland and Pammer, 2010; Holmqvist *et al.*, 2011). Hence the results should be supported by other techniques.

### *7.1.2 Variety of eye-tracking metrics*

The eye-tracking tools collect a wide variety of different metrics that focus both on fixations and saccadic movements (transitions between fixations). These metrics include fixation count and duration, time to first fixation, scan paths, pupil sizes and many more. Holmqvist *et al.* (2011) defines fixation as "*a time remains still over a period of time"* (p.21). The metrics related to fixation count and duration are the most popular metrics to use in modern studies (Jacob and Karn, 2003; Holmqvist *et al.*, 2011; Kapitaniak *et al.*, 2015). As the eye-tracking body of knowledge is almost non-existent in rail industry, it was decided to use the most popular metrics from other disciplines. This allows using data from other industries as a starting point for designing the experiment. According to previous research, all of the metrics provide information on cognitive processes.

Number of fixations, or total fixation count (TFC), is a sum of fixations recorded on the area of interest (AOI) in a set period of time. High number of fixations can suggest semantic importance or informativeness of the AOI (Yarbus, 1967; Henderson and Hollingworth, 1999; Jacob and Karn, 2003; Poole *et al.*, 2004). High TFC can also be a sign of difficulty in interpreting information (Ehmke and Wilson, 2007; Holmqvist *et al.*, 2011), complexity (Dzeng *et al.*, 2016) and poor search efficiency (Goldberg and Kotval, 1999). Experience was found to be important factor affecting fixation count statistics. People with previous experience in the area under investigation tend to have less fixations (Megaw and Richardson, 1979; Reingold *et al.*, 2001). However, there are studies showing that more experienced pilots have more fixations in a cockpit than novice pilots (Kasarkis *et al.*, 2001). This does not contradict the previous theories about the connection between experience and a number of fixation though. It only shows that in some domains more experienced users employ more optimised strategies and reduce a number of unrequired fixations in non-critical areas or increase it in critical areas to double check the readings.

Holmqvist *et al.* (2011) defines fixation duration as "*a period of time when the eye is relatively still*" (p.377). In reality, a human eye is rarely completely still and fixation durations are calculated by the fixation detection algorithm provided by developers of eye-tracking equipment. In this experiment Tobii I-VT fixation filter was used (Olsen, 2012). It was responsible for detecting fixations and subsequent durations of those. Total fixation duration (TFD), also known as total visit duration or total dwell time, is a frequently used metric in eye-tracking research. Higher TFD on an object indicates object's higher informativeness and participant's interest in the object (Holmqvist *et al.*, 2011). It also can indicate poorer situational awareness and uncertainty (Ottati *et al.*, 1999). In usability studies, longer dwell time suggest difficulties with information extraction (Goldberg and Kotval, 1999). Similarly, research in reading domain shows that high TFD are indicative of more complex and less frequent words (Rayner, 1998). Finally, an upcoming conscious choice is indicated by longer dwell times as people tend to fixate longer on something they going to pick from other objects (Shimojo *et al.*, 2003). Hauland (2008) argues that dwell time is much better unit to investigate situational awareness as it represents cognitive meaning compared to a single fixation.

Average fixation duration (AFD) for an area of interest is equal to TFD divided by TFC. As with other metrics, AFD statistics can indicate certain cognitive processes. Longer durations is a sign of deeper processing (Holmqvist *et al.*, 2011), criticality of elements (Shinar *et al.*, 1977; Harris and Christhilf, 1980), a sign of issues with information extraction (Holmqvist *et al.*, 2011). At the same time, shorter fixations also indicate issues with information extraction due to higher mental workload and stress (Miura, 1990; Unema and Rotting, 1990; Van Orden *et al.*, 2001), and usability problems. Holmqvist *et al.* (2011) makes a distinction between higher workload associated with higher AFD (which a human can complete without problems) and with lower AFD (causing performance issues due to additional stress).

Researchers differentiate between fixations hence averaging fixation duration could mean mixing several types of cognitive processing together (Henderson *et al.*, 1999; Holmqvist *et al.*, 2011). Moreover, AFD was found to be the most affected by an individual. The AFD tends to remain similar in a repeat task (Andrews and Coppola, 1999). Hence statistical analysis of the samples is more important than averaging across the samples. On the other hand, in the case of matched trials averaging can be sufficient as endogenous component is constant across the trials. The eye gaze analysis software used in the trial does not account for the recent discoveries on a lag between visual attention and a fixation which could be as big as ¼ of a second (Deubel, 2008; Holmqvist *et al.*, 2011), where visual attention is moving faster. This lag should not however massively skew data for 10/15 seconds average and can only affect fixations recorded in the last 250ms of a timeframe. Watching recordings demonstrated that drivers tend to shift their gaze to in-cab (speed, door controls) right after full stop or departure. As the study is not analysing sequences and patterns of fixations, it is assumed that fixations equal cognitive processing.

## 7.2 Methodology

Section 7.1.2 reports a strong connection between eye-tracking metrics and workload and stress. Lindner (2013) claims that infrastructural parameters and constraints can influence a set of PSFs contributing to human errors. Hence contributions to workload/stress from the physical environment can be studied through the eye-tracking research. This can be done by direct comparison between locations with different design features.

The results of the previous chapters propose many areas for the eye-tracking investigation. However, many of those had to be excluded due to limitations presented by available eye-tracking equipment, e.g. underground operations. A number of trial runs conducted before the study showed that low resolution (640x480) and poor camera quality does not allow good data collection in those parts of the system. Hence research locations were chosen based on covering as many design features highlighted in Chapter 5 and Chapter 6 as possible and fitting with eye-tracker parameters.

### *7.2.1 Stations*

Four stations have been selected for the study. Those are Pelaw, Heworth, Felling and Gateshead Stadium. Twenty runs were conducted between February and June 2015. All of the runs were in the same direction from Pelaw to Gateshead Stadium. Figure 11Figure 14 below provide outline schemes of the stations under investigation. The diagrams are not to scale.

The exploratory character of the study meant that the stations were selected on the basis of:

- Difference in patronage levels (Heworth and Pelaw – the highest, Felling – the lowest),
- A mix of built-over and overground stations,
- A mix of open and confined layouts,
- A mix station types,
- Different locations of a mirror and a running signal in respect to a stopping position,
- Different location of a stopping position in respect to a platform’s edge.

Such variety of station designs should be sufficient to explore drivers’ performance in different station layouts and draw some hypotheses for further investigation.

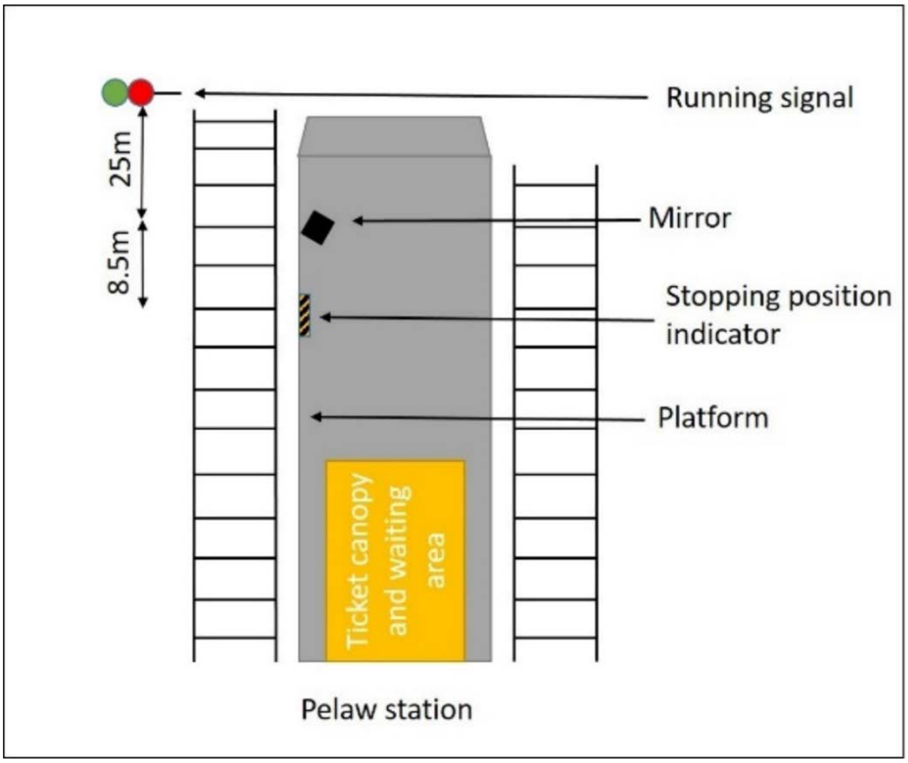


Figure 11. Layout of Pelaw station (Type 2)

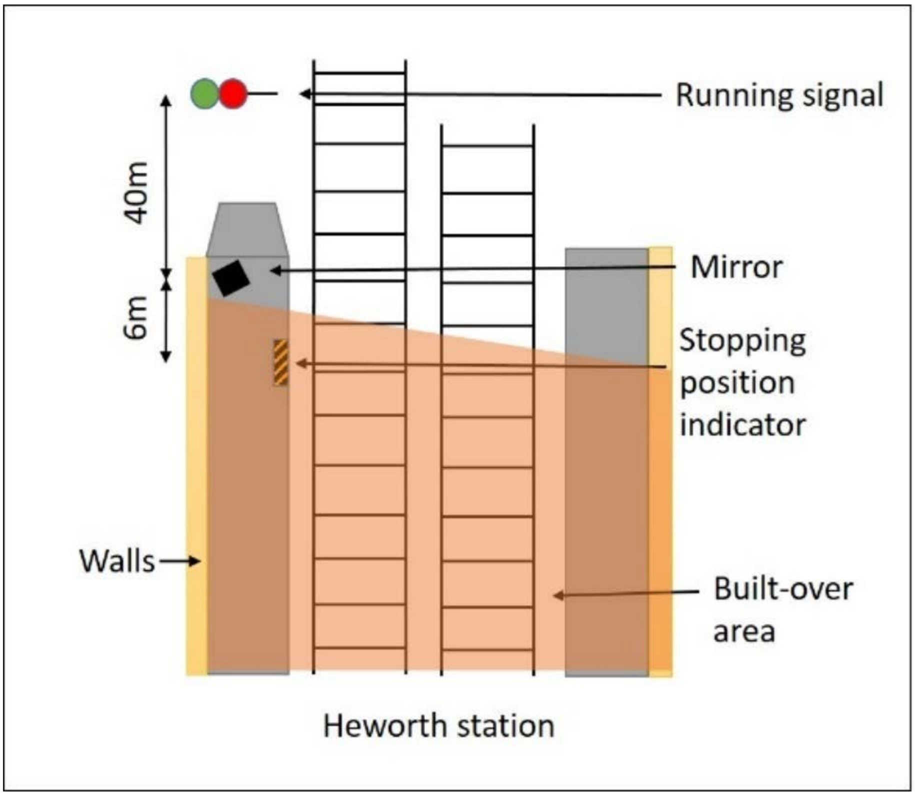


Figure 12. Layout of Heworth station (Type 1)

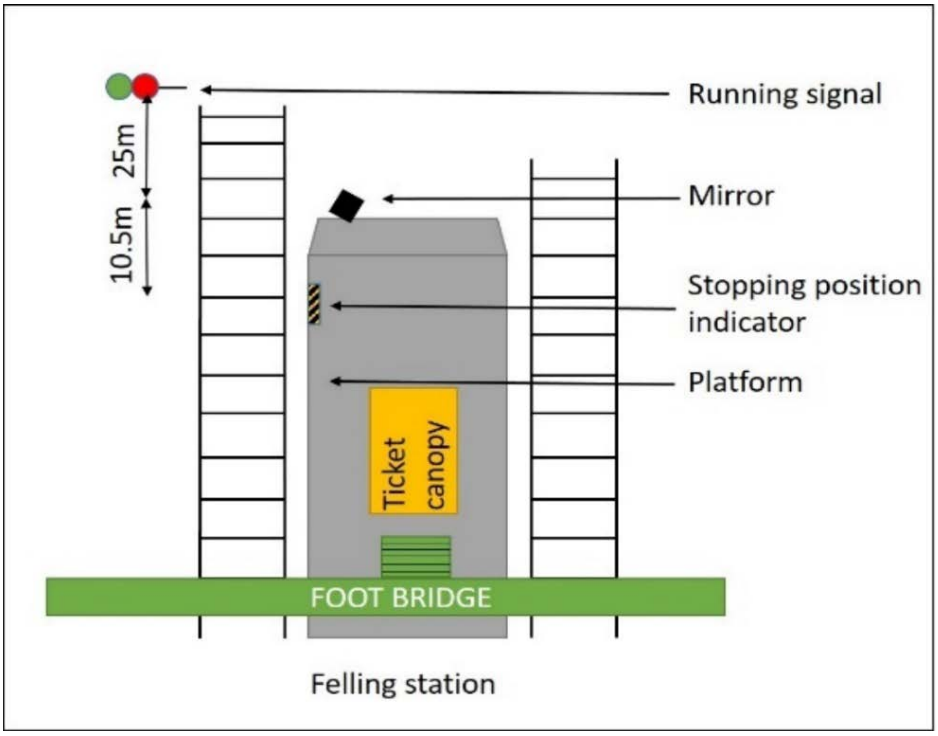


Figure 13. Layout of Felling station (Type 2)

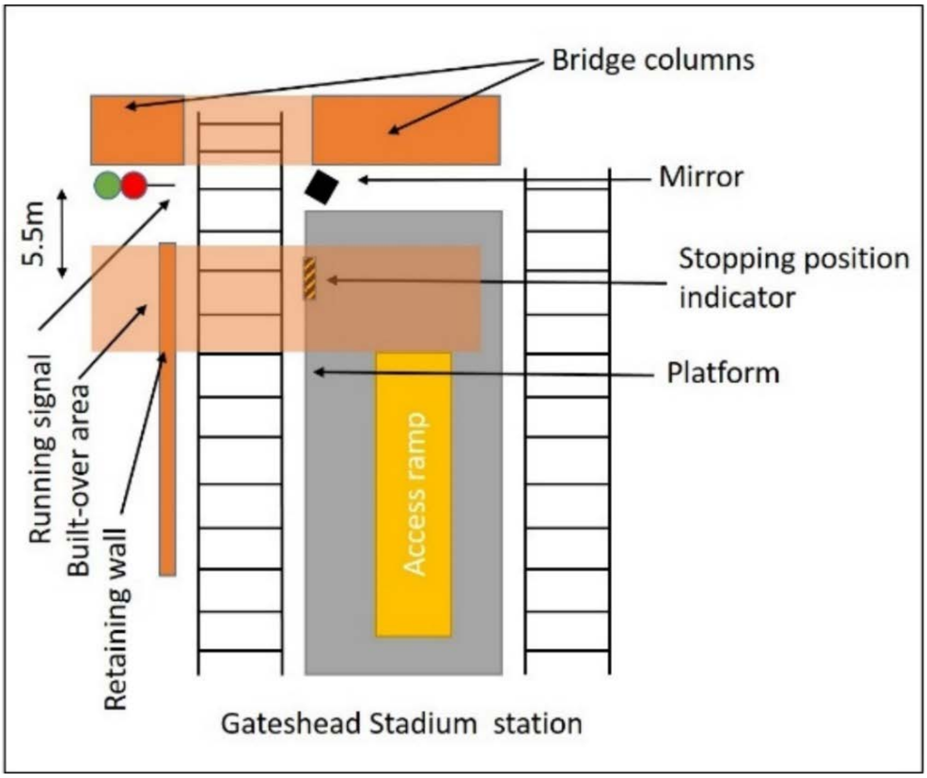


Figure 14. Layout of Gateshead Stadium station (Type 2)

### *7.2.2 Participants*

Four drivers participated in the study. All of the participants are drivers responsible for training and were selected for their experience. This was done to avoid differences in gaze behaviour based on experience and route knowledge. Personal factors such as various health conditions and dysfunctions (Coeckelbergh *et al.*, 2002), and age (Ho *et al.*, 2001) can also affect fixation patterns. These factors were also taken into account during a candidate selection process.

### *7.2.3 Areas of interest*

Four elements of stations' physical environment, which are present at all four of stations under investigation, are selected as areas of interest (AOI). Those are:

- Mirrors,
- Stopping position markers,
- Platforms,
- Running signals

For departures only the mirrors and the signals are analysed as those are the only elements visible to a driver. Holmqvist *et al.* (2011) report that the best practice is to have margins between AOIs and no overlapping AOIs. The latter is achieved by carefully specifying AOIs in the analysis software but the former cannot be controlled in in-field experiment. The raw data is analysed in Tobii Studio software.

The timeframe analysed with respect to these AOIs is 15 seconds before a complete station stop. A preliminary investigation showed that 15 seconds is approximately the time from station entry to the complete stop. Moreover, 15 seconds intervals were used by RSSB (2005) before. As in many locations drivers approach a station from a curve, it is assumed that participants do not focus on the AOIs prior to this timeframe. For the departures 10 seconds timeframe was selected after consultations with drivers. The cut-off point used for departures is a moment when a train starts moving.

When station approaches are considered, the AOIs under investigation are dynamic. In other words, the position of the AOIs in respect to a driver's position changes as a train drives forward. As the equipment and software do not have automatic tracking of dynamic AOIs, such AOIs can be manually adjusted and analysed (Holmqvist *et al.*, 2011). To do so, the 15-seconds timeframe is split into 3-seconds timeframes where positions of AOI are adjusted to reflect the change in relative position. Some AOIs are not defined until the second or third timeframe as their size is too small to create any meaningful statistics. The TFD measure is the most resilient to the issues caused by the chosen analysis method as it is rarely affected by splitting the approach timeframe. Such split can create double accounting of fixations (affecting AFD) but not TFDs.

According to Kapitaniak *et al.* (2015) conspicuity (ability to detect objects) is affected by instructions and search targets. The searched object is 3 times more visible than an object observed independently (Bremond, 2000). On the other hand, the research by Bremond (2000) focused on car driving and roadside environment where visual clutter is higher and this can affect gaze performance (Ho *et al.*, 2001). The AOIs chosen for this study form physical environment required to carry out arrival and departure tasks safely. Hence there should be no discrepancy between the AOIs in terms of conspicuity.

### *7.2.4 Metrics*

As shown in Section 7.1.2 there is a wide variety of eye-tracking metrics. Some of those can suggest completely opposite trends based on the literature review. On the other hand, one can note that some measures indicate similar things. High TFC, similarly to high AFD, can be a sign of importance or criticality of an object, higher semantic value. High TFC can also indicate difficulties in interaction with physical environment and information extraction, and inefficient search strategies. These usually correlate with additional workload and stress thus causing lower AFD. According to Kapitaniak *et al.* (2015) "*overall increase in requirements and complexity of the task tends to reduce AFD and increase TFC*" (p.950). Crundall *et al.* (1998) conducted a series of eye-tracking experiments on different types of roads to show that a busy sub-urban road (going through a village) leads to a higher sampling rate (high TFC/low AFD), compared to empty rural roads. They assumed, based on previous research, that driving on such suburban roads would increase task demands due to a visual clutter, and subsequently increase cognitive demands. Crundall *et al.* (1998) also referred to the research by Beck (1985), who showed that anxiety in people leads to more active search strategies. The higher sampling rate can be considered a part of such strategies.

A different experiment with a group of trainee anaesthetists, Schulz *et al.* (2011) used a simulated critical incident to increase mental workload. The results showed that fixation durations rise in a routine part of a surgery but start decreasing as an unexpected event progresses. Such events are associated with the highest mental workload experienced by participants.

Based on the above research, drop in TFC and an increase in AFD should indicate usability improvements. Lower TFC/higher AFD has been proven to be an indicator of expertise. Reingold and Charness (2005) showed that expert chess players require fewer fixations of a similar length than novice players. Savelsbergh *et al.* (2002) also demonstrated that more experienced football goalkeepers employ search strategies with fewer but longer fixations before a penalty kick. Both studies suggest that previous experience helps in developing more sophisticated search strategies where the same amount of information is extracted from fewer AOIs (a narrower field of search). This can be inferred as supporting reduced workload in the low TFC/high AFD, assuming that more sophisticated and efficient search strategies should reduce anxiety and workload. In order to avoid the effects of previous expertise, only experienced driver coaches were selected for the trials.

### *7.2.5 Eye-tracker*

According to Hessels *et al.* (2015), Tobii manufacture the most popular eye-tracking devices in Northwest Europe. "*Tobii Glasses*" (first generation) eye-tracker was used in the experiment. Figure 15 below shows one of the participants wearing the set. The eye-tracking set consists of eye-tracking glasses and a recording device.

**Figure 15. One of the participants wearing Tobii glasses in a metrocar cab**

The set has 30 hz sampling frequency and uses a 9-point calibration algorithm which is performed before each trial. The sampling frequency provided by the set is not great, compared to 60+ Hz eye-trackers available on the market. Andersson *et al.* (2010) demonstrated a reverse relationship between a sampling frequency and a probability of a measurement error. With a frame produced every 33ms, all of the differences below this parameter should be treated with caution.

### *7.2.6 Various samples and data quality*

Even though in-field eye-tracking studies can provide rich data sets, those also suffer from limited controls (De Ceunynck *et al.*, 2015). The trials were designed in a way that tried to exclude as many uncertainties as possible. The participants were of similar experience, ethnical background, gender and had good eye vision. Even then, out of 26 trials only 20 were considered successful upon preliminary inspection of recordings. Factors that influenced exclusion of trials were direct sunshine and drivers adjusting position of the eye-tracker (without re-calibration) during the experiment. According to Holmqvist *et al.* (2011), loss of up to 32% of the participants had been reported previously thus losing 20% of the participants can be considered a good result. Moreover, Sodhi *et al.* (2002) lost 80% of the trials due to data quality but the study used even more inferior eye-tracking devices. However, there were concerns with data quality of the retained experiments due to two reasons.

Firstly, the trials returned different sample retention rate. This measure is calculated by Tobii Studio software and means percentage of frames recorded with eye-tracking data. Three trials were found to have this measure of less than 50%. RSSB (2005) had to exclude more 40% of the sample in order to achieve 84% sample retention rate. It is important to note that due to risks associated with field trials, a terminus to terminus journey was recorded and this metric is presented for the entire recording, from the beginning of the trial (at South Hylton or South Shields) to Gateshead Stadium part. Secondly, the starting terminus for the trials was different meaning that some drivers were driving for 30 minutes before reaching the section of the system under investigation whereas other drivers only for one minute.

The statistical analysis showed that neither of these two factors influenced results as different samples (excluding experiments affected by these factors) trialled against the whole sample did not show any statistical difference. Moreover, no statistical relationship between the variables changed after reducing a sample size. Hence it was decided to present only results for the entire sample (20 runs).

Upon visual inspection of the recordings, it was noted that some of the eye fixation marker is clearly offset from fixation locations. According to Holmqvist *et al.* (2011), it is acceptable to repair this data by offsetting the AOIs in a similar way. The offset was found using departure recordings, as the participants are stationary and relative position is constant.

### 7.2.7 Analysis

As the data for station arrivals was split into five smaller timeframes, it is first summed up to produce statistics for the 15-seconds timeframes. The datasets containing 48 and 24 variables are created for arrivals and departures respectively.

The variety of station layouts allow comparison of the same AOIs at different stations. This is done in two stages. Firstly, the data is explored using descriptive statistics for all metrics. Next, the uncovered relationships are tested using Paired Samples T-test or Wilcoxon Signed Ranks Test. The tests are selected based on normality of the variables. According to Moore *et al.* (2009) "*sample size strongly influences the P-value of a test*" (p.465) when it comes to comparison of means. They claim that effects with p-values slightly larger than widely acknowledged 0.05 can become statistically significant (at 95% confidence interval) in larger samples. Considering the small sample of trials (N=20) and even smaller sample of participants (N=4), it is necessary to report p-value of up to 0.1 to minimise probability of Type II error (accepting null hypothesis when it is false). Even studies with bigger samples reported findings on *p*-level of 0.1 (Aitchinson and Davies, 2009). Tripathi and Borrion (2016) demonstrated through sensitivity analysis how increase in sample size can make affect the null hypothesis.

Additional statistical tests (Kruskal-Wallis H test) are performed using secondary factors recorded during the experiments. Those factors are weather and individual participant's ID. The weather is presented by three categories: rainy, cloudy and sunny. The ID allows testing whether personal factors affect gaze performance even when participants are carefully selected to provide uniformity of a sample. Naghiyev *et al.* (2014) showed that train drivers' visual strategies are not uniform. However, the authors do not mention whether any controls were applied to provide uniformity of the participant pool.

## 7.3 Results

### *7.3.1 Arrivals for all elements under investigation*

According to Figure 16 the participants spend more time fixating on the AOIs at Heworth and Gateshead stadium. As it is demonstrated further in this chapter, the most difference arises from TFD on platforms. Other elements of physical environment (in-cab, scenery and track) potentially detract from the AOIs too.

When paired samples T-Test (all variables are normally distributed) is performed (Table 20), the only statistically significant difference is found between Felling and Heworth. Another two differences (Felling – Gateshead Stadium and Heworth – Pelaw) are close to statistical significance with *p-value<.085*. The results of the t-test show that the drivers definitely spend least time fixating on the AOIs at Felling and the most time at Heworth.

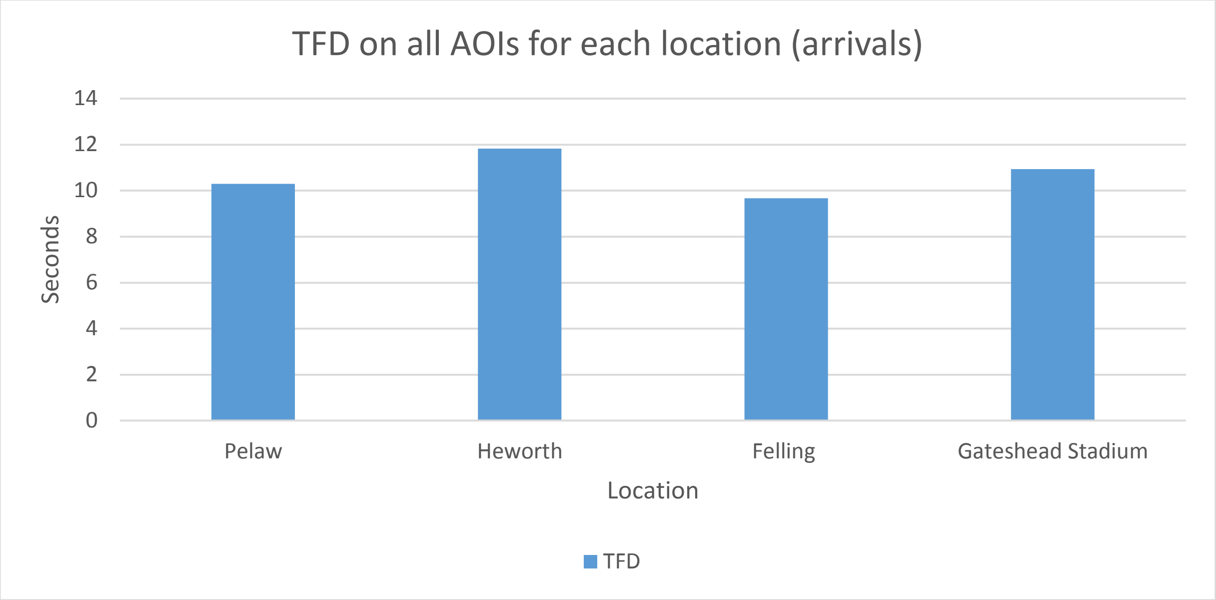


**Figure 16. TFD for all AOIs at each station (arrivals)**

| | **Pelaw** | **Heworth** | **Felling** |
|---|---|---|---|
| **Heworth** | 1.828* | | |
| **Felling** | -1.104 | -2.647** | |
| **G. Stad** | 1.157 | -.986 | -2.005* |

**Table 20 Paired samples T-Test result for TFD on all AOIs (arrivals). *p<0.1, **p<0.05**

### 7.3.2 Mirrors during arrivals

Figures 17-19 below provide a summary of TFD, TFC and AFD statistics for the mirrors across all locations. Drivers spend less time looking on Heworth mirror compared to other stations. Gateshead Stadium mirror has slightly higher TFD than other locations. Pelaw and Felling show relative reverse proportionality for TFC and AFD statistics, effectively showing higher TFC and lower AFD than other stations. This is more pronounced for Pelaw than for Felling, both Type 2 stations.

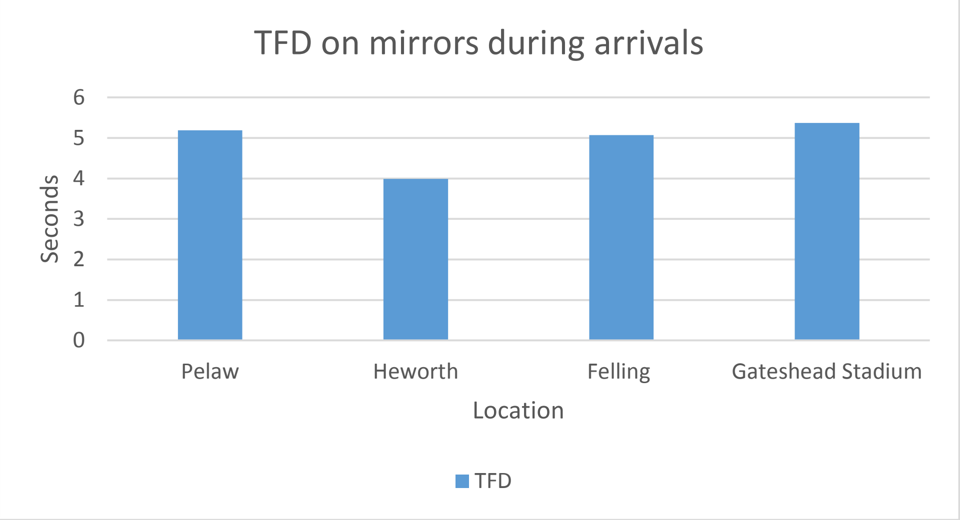


Figure 17. TFD on mirrors during arrivals

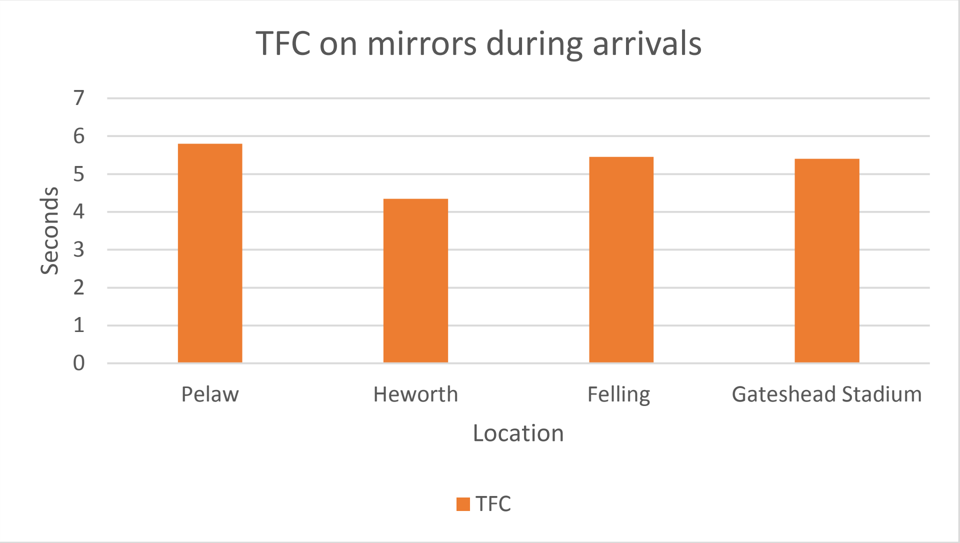


Figure 18. TFC on mirrors during arrivals

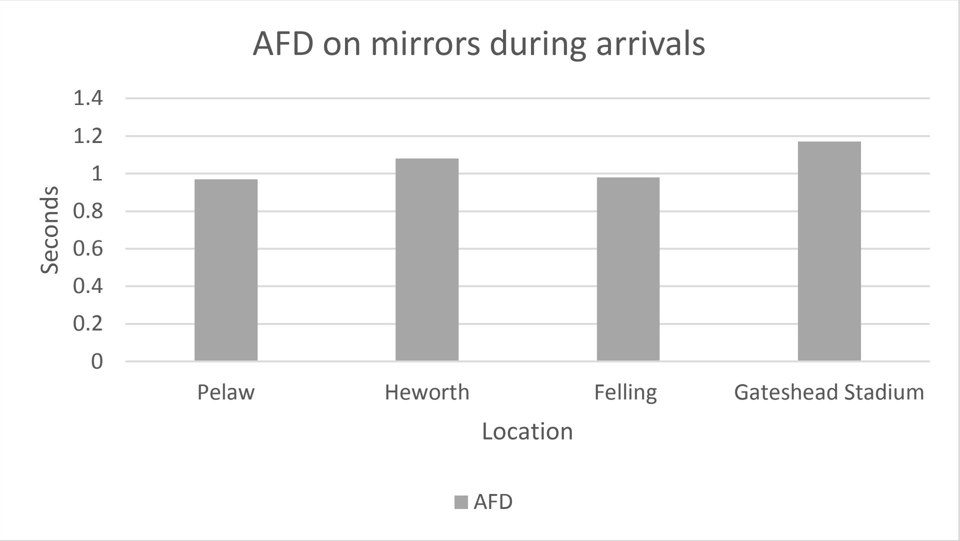


Figure 19. AFD on mirrors during arrivals

### *7.3.3 Platforms during arrivals*

Figures 20-22 below provide a summary of TFD, TFC and AFD statistics for the platforms across the locations. A clear difference is demonstrated in terms of TFD measures. Heworth and Gateshead have considerably higher gaze durations on the platforms than Pelaw and Felling. Relative proportionality is shown by Heworth platform having both the highest TFC and AFD. The platform at Felling receives the least fixations but has higher AFD than Gateshead Stadium and Pelaw. Compared to Felling, Pelaw has higher TFC but marginally lower AFD.

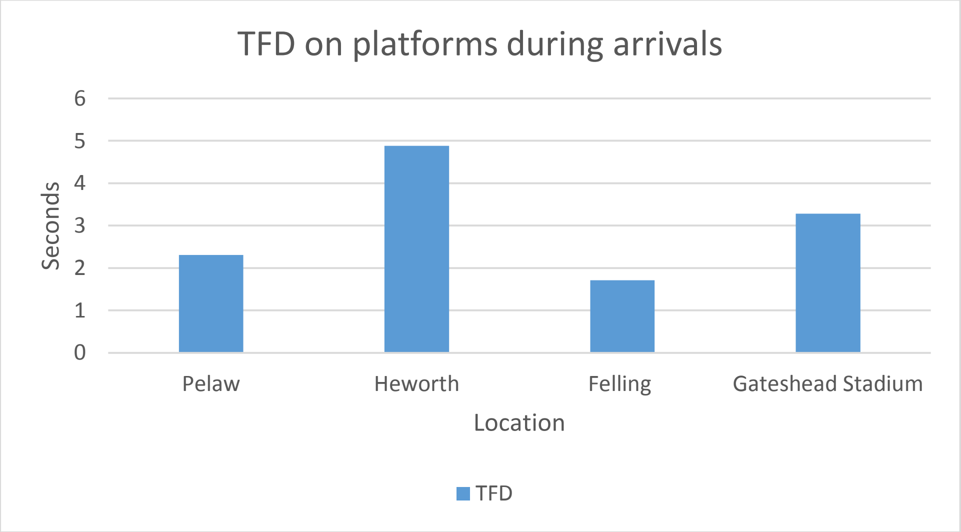


Figure 20. TFD on platforms during arrivals

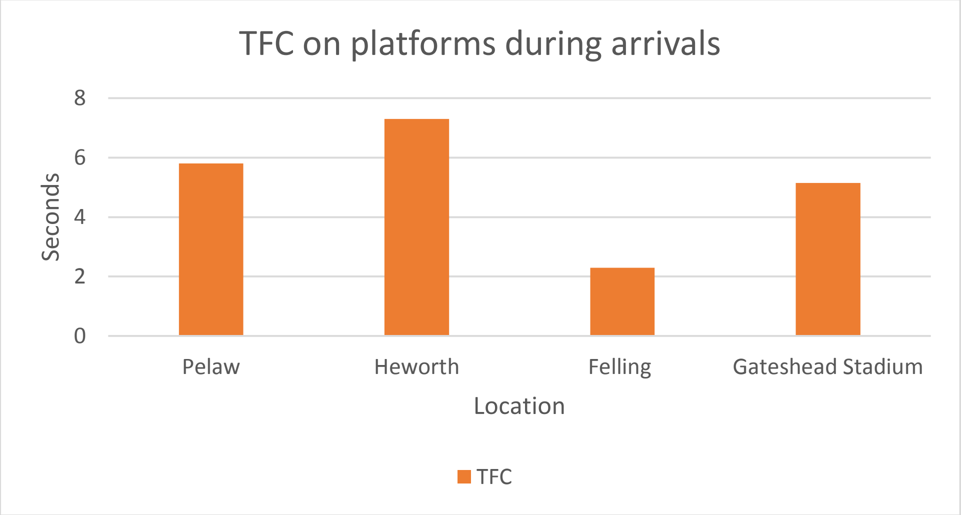


Figure 21. TFC on platforms during arrivals

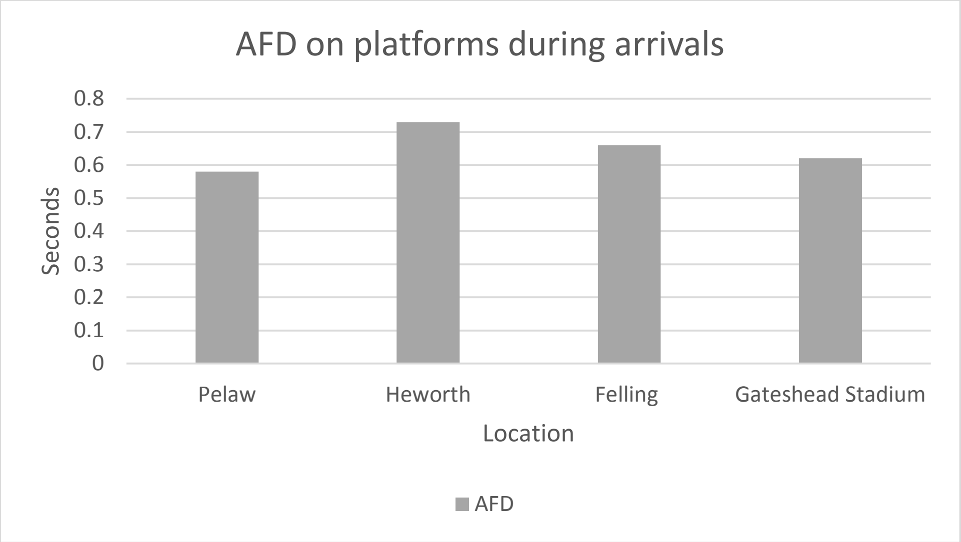


Figure 22. AFD on platforms during arrivals

### *7.3.4 Signals during arrivals*

Figures 23-25 below provide a summary of TFD, TFC and AFD statistics for the running signals under investigation. The participants spend considerably less time looking on the signal at Gateshead Stadium compared to other stations. The AFD and TFC statistics is mostly proportional but Heworth shows higher TFC and lower AFD compared to Felling.

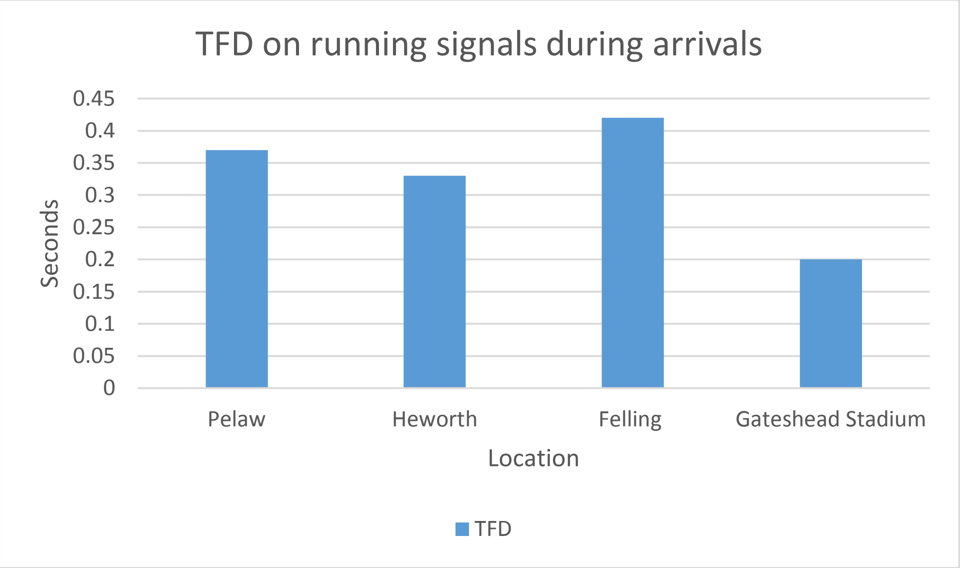


Figure 23. TFD on running signals during arrivals

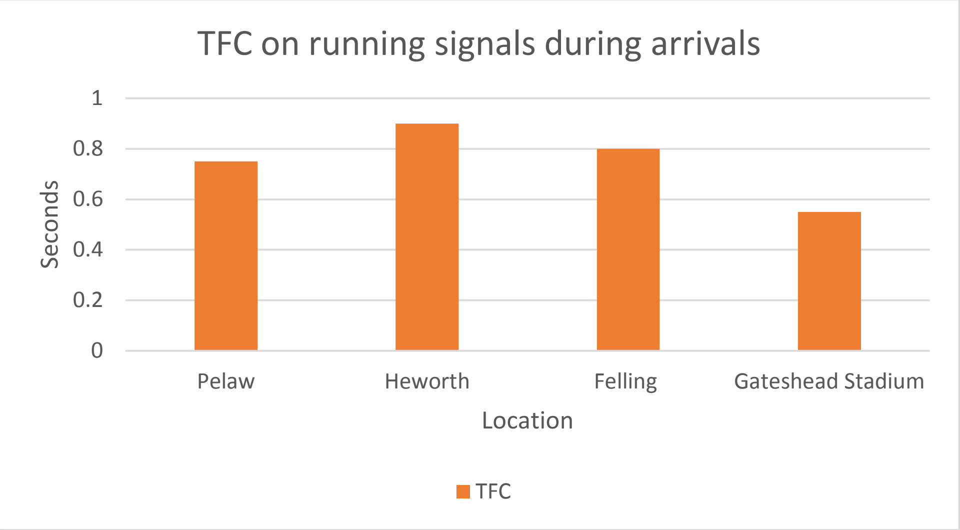


Figure 24. TFC on running signals during arrivals

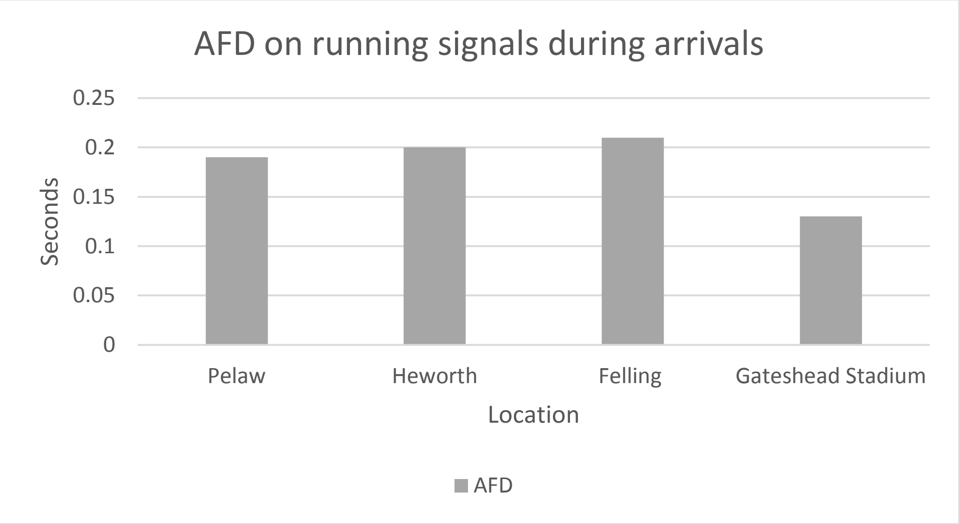


Figure 25. AFD on running signals during arrivals

### *7.3.5 Stopping position indicators during arrivals*

Figures 26-28 below provide a summary of TFD, TFC and AFD statistics for the stopping position indicators under investigation. Heworth shows the highest TFD on this element whereas Gateshead Stadium has the lowest. Average fixation durations have only minor differences for all stations but Gateshead Stadium, where approximately 50% decrease in this metric can be observed. Moreover, the station demonstrates the highest TFC on the stopping position indicator among all locations.

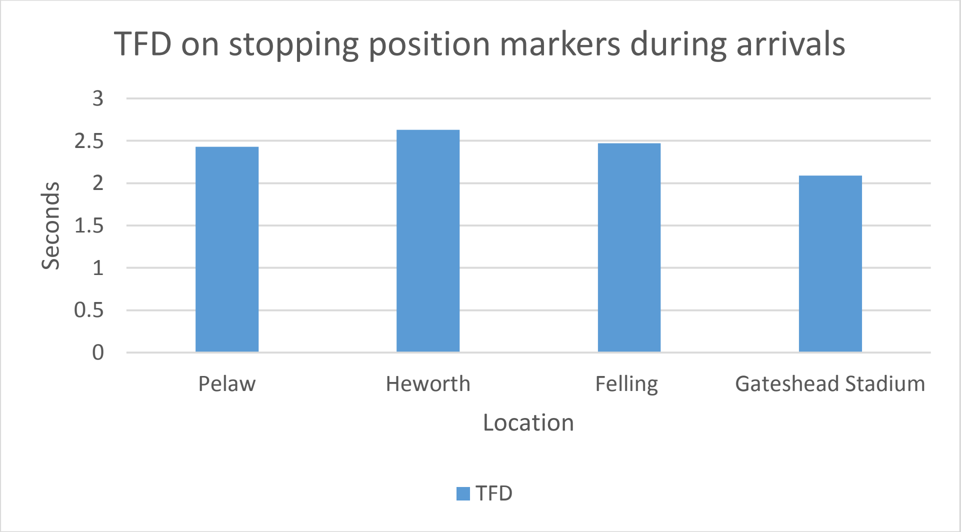


Figure 26. TFD on stopping position markers during arrivals

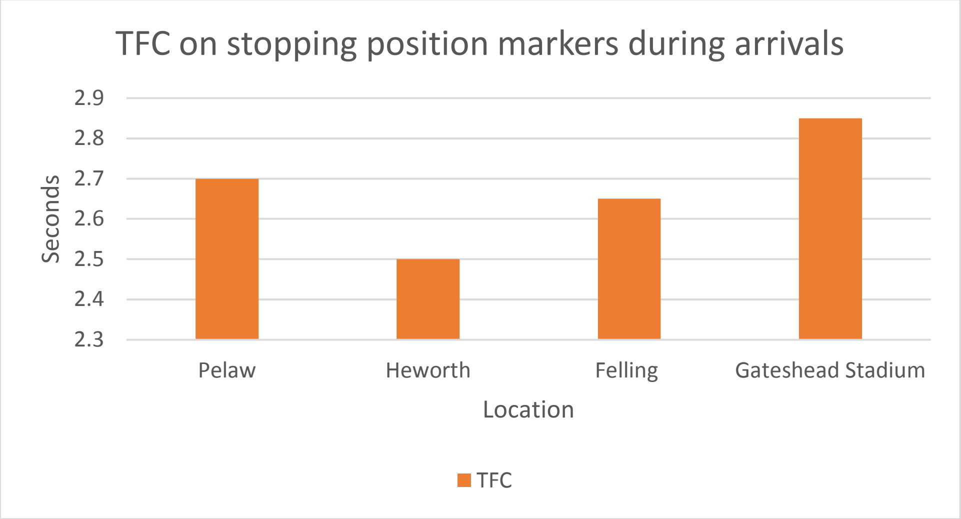


Figure 27. TFC on stopping position markers during arrivals

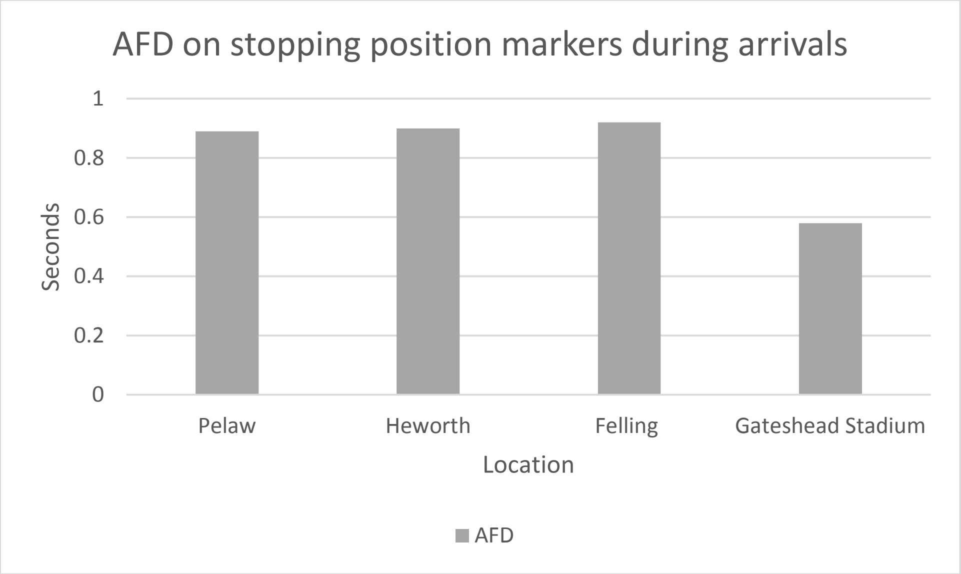


Figure 28. AFD on stopping position markers during arrivals

### 7.3.6 *Wilcoxon Signed Rank Test for AOIs during arrivals*

As Table 21 below shows, only some of the arrival variables showed any statistically significant difference between the locations. Out of 24 relationships, 5 are statistically different for TFD, 5 – for TFC and 1 – for AFD at 95% confidence interval. When it is reduced to 90%, another 6 differences are found. In terms of elements, the majority of differences are found for platforms (TFC and TFD). AFD on Gateshead Stadium signal is different to signals at all other stations. Heworth mirror demonstrates statistically significant difference with all other mirrors in terms of total gaze duration on it. The signals do not demonstrate any difference between locations.

| | | Metric | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | TFD | | | TFC | | | AFD | | |
| | | Pelaw | Heworth | Felling | Pelaw | Heworth | Felling | Pelaw | Heworth | Felling |
| **Mirror** | Heworth | -1.979** | | | -2.235** | | | -1.008 | | |
| | Felling | -0.093 | -1.717* | | -.699 | -1.463 | | -.243 | -.224 | |
| | G.Stad | -.448 | -1.867* | -.635 | -.549 | -1.225 | -.048 | -1.941* | -.597 | -1.792* |
| **Platform** | Heworth | -2.856** | | | -2.639** | | | -1.419 | | |
| | Felling | -1.494 | -3.173** | | -1.820* | -3.685** | | -.560 | -.560 | |
| | G.Stad | -2.501** | -1.568 | -2.333** | -2.803** | -1.431 | -2.991** | -.597 | -1.456 | -.560 |
| **Signal** | Heworth | -.157 | | | -.206 | | | -.235 | | |
| | Felling | -.078 | -.157 | | -.420 | -.233 | | -.471 | -.408 | |
| | G.Stad | -1.479 | -.756 | -.863 | -.863 | -.511 | -1.026 | -1.071 | -.979 | -1.02 |
| **Stop.Pos.** | Heworth | -.871 | | | -.477 | | | -.457 | | |
| | Felling | -.149 | -.037 | | -.263 | -.177 | | -.336 | -.149 | |
| | G.Stad | -.610 | -1.112 | -1.408 | -.257 | -.810 | -.317 | -1.917* | -2.343** | -1.730* |

**Table 21. Difference in gaze behaviour at different locations (arrivals). *p<0.1, **p<0.05**

### 7.3.7 *Departures for all elements under investigation*

Figure 29 below shows TFD on all for elements at each location. Heworth shows the highest TFD with almost 80% of the analysed departure time spent fixating on either the mirror or the signal. In terms of the lowest TFD, Felling and Pelaw show comparable results. The TFD value at Gateshead Stadium is between the lowest and the highest results.

Results of the Wilcoxon Signed Rank test (Table 22) confirm the lowest TFD at Felling. However, there is no statistically significant difference between Heworth and other stations at even *p<0.1*. The difference between Heworth and Pelaw shows p-value close to 0.1 but still higher.

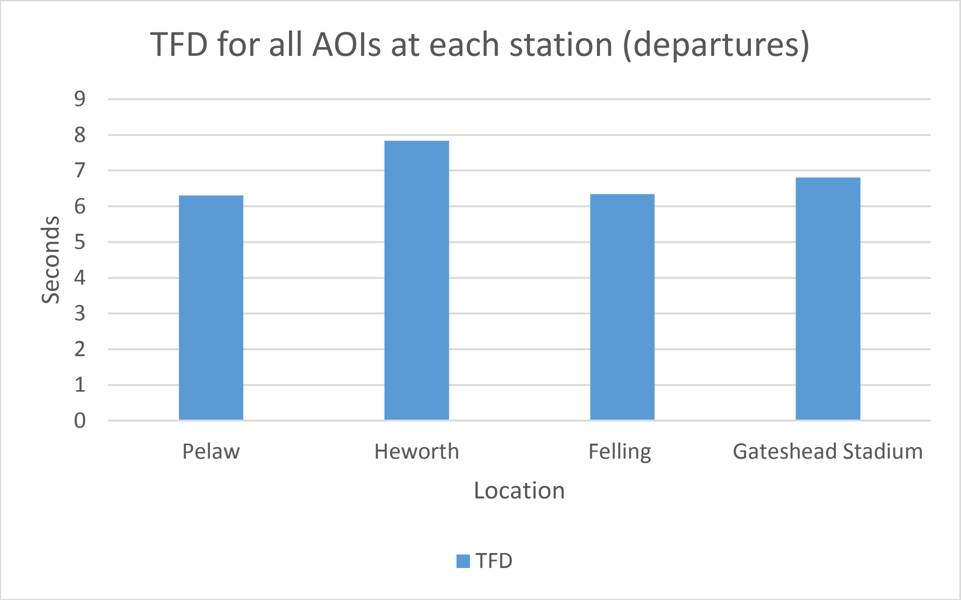


**Figure 29. TFD for all AOIs at each station (departures)**

| | **Pelaw** | **Heworth** | **Felling** |
|---|---|---|---|
| **Heworth** | -1.605 | | |
| **Felling** | -1.120 | -2.315** | |
| **G. Stad** | -1.419 | -.672 | -1.661* |

**Table 22. Wilcoxon Signed Ranks test result for TFD on all AOIs (departures). *p<0.1, **p<0.05**

### *7.3.8 Mirrors during departures*

Figures 30-32 below summarise TFD, TFC and AFD statistics for the mirrors on departures. The participants spend the most time looking on the mirror at Heworth, and the least looking on the mirror at Felling. Pelaw and Gateshead Stadium mirrors had TFD in-between the highest and the lowest results, with the mirror at Gateshead Stadium showing slightly higher TFD. The TFC/AFD statistics is proportional at all stations but Pelaw, where the second highest TFC and the lowest AFD numbers are observed.

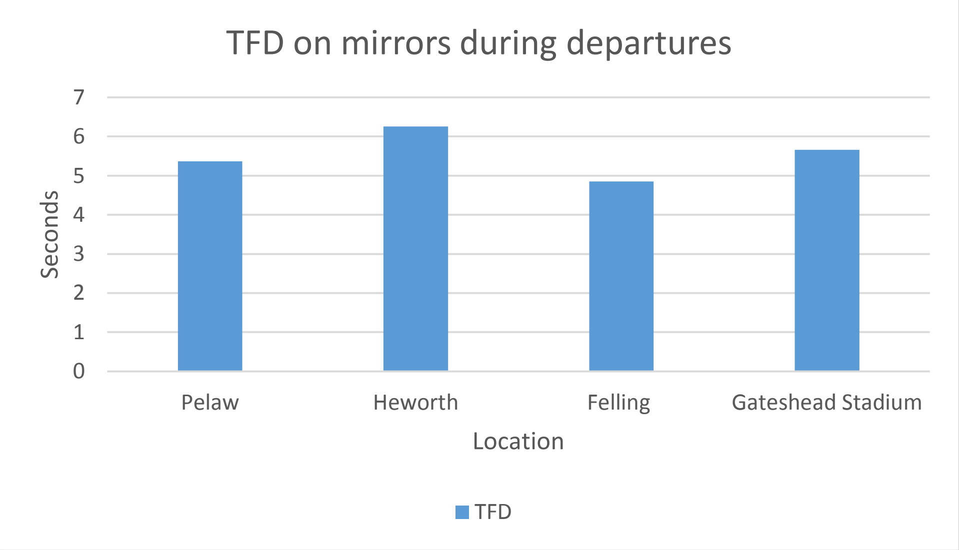


**Figure 30. TFD on mirrors during departures**

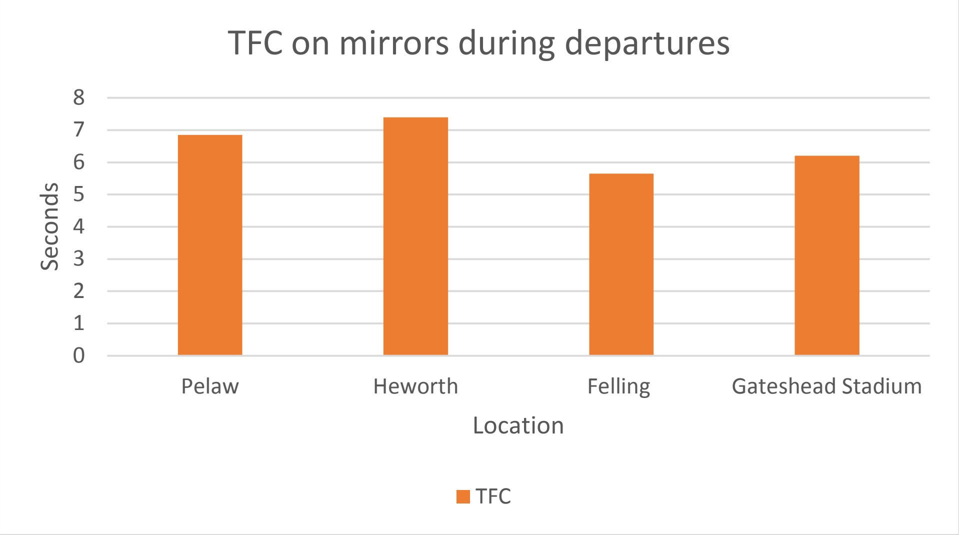


**Figure 31. TFC on mirrors during departures**

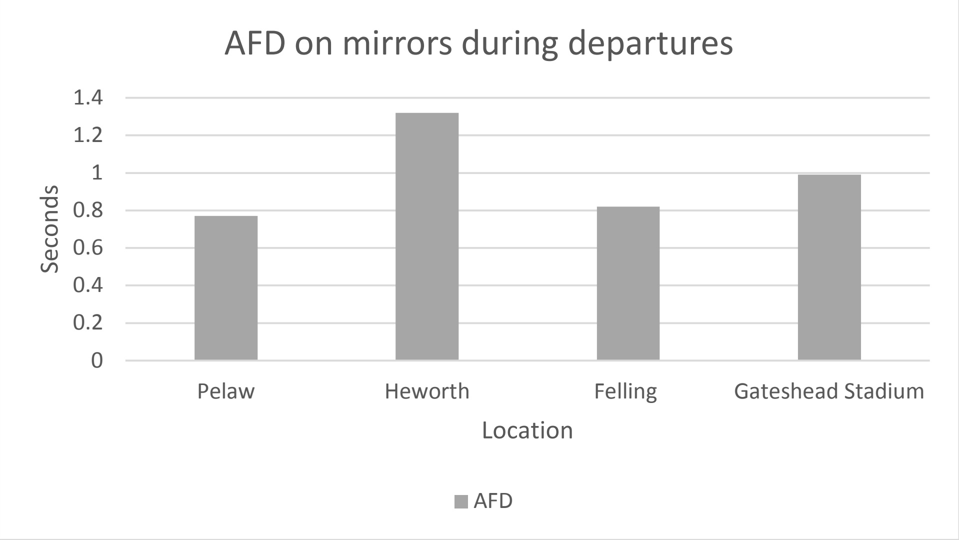


**Figure 32. AFD on mirrors during departures**

### *7.3.9 Signals during departures*

Figures 33-35 below summarise TFD, TFC and AFD statistics for the mirrors across all locations. In terms of TFD, the signal at Pelaw shows the lowest values with the signal at Heworth showing almost double of it. The signal at Felling attracts almost similar amount of gaze time to Heworth signal. On contrary to the mirrors results, reverse proportionality is observed for all stations but Pelaw, where both metrics (TFC/AFD) are the lowest in the sample. Heworth signal demonstrates higher TFC/lower AFD scenario. Gateshead Stadium signal shows an opposite trend.

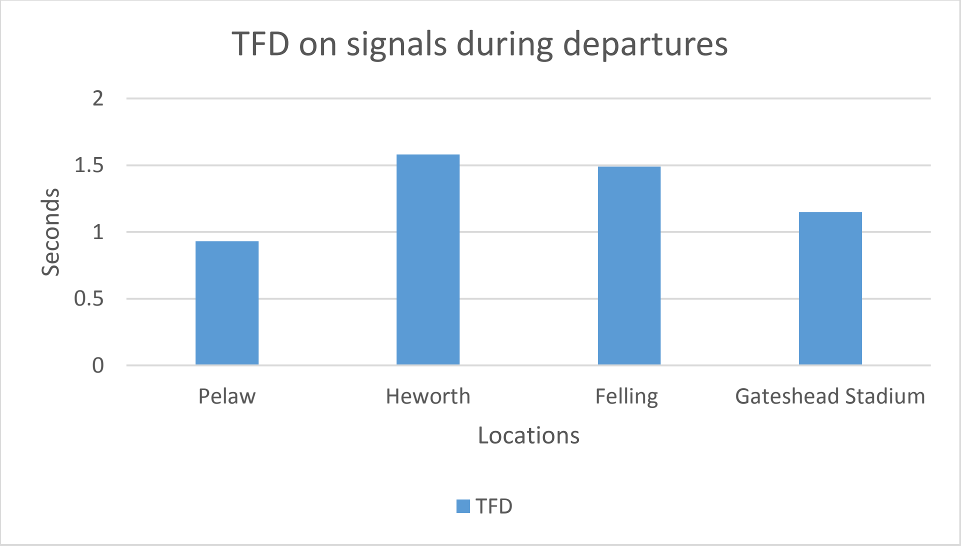


Figure 33. TFD on signals during departures

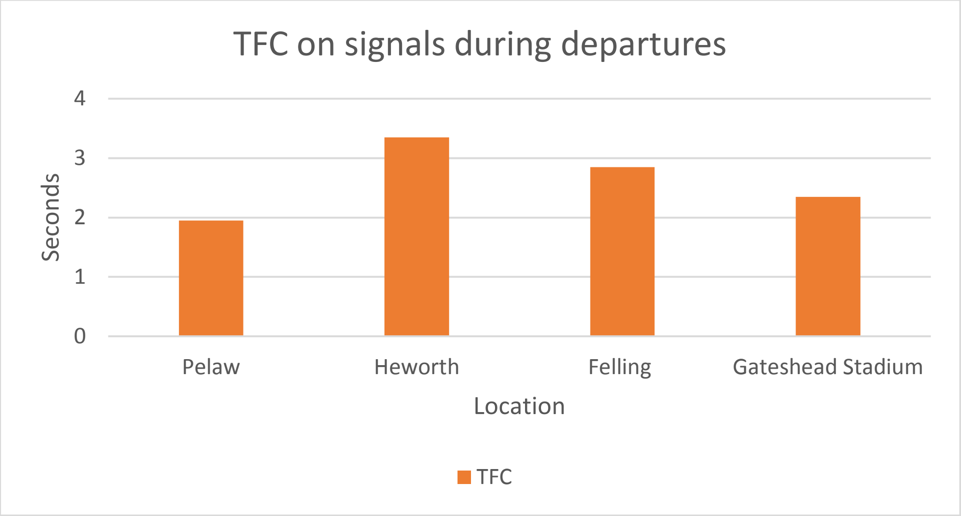


Figure 34. TFC on signals during departures

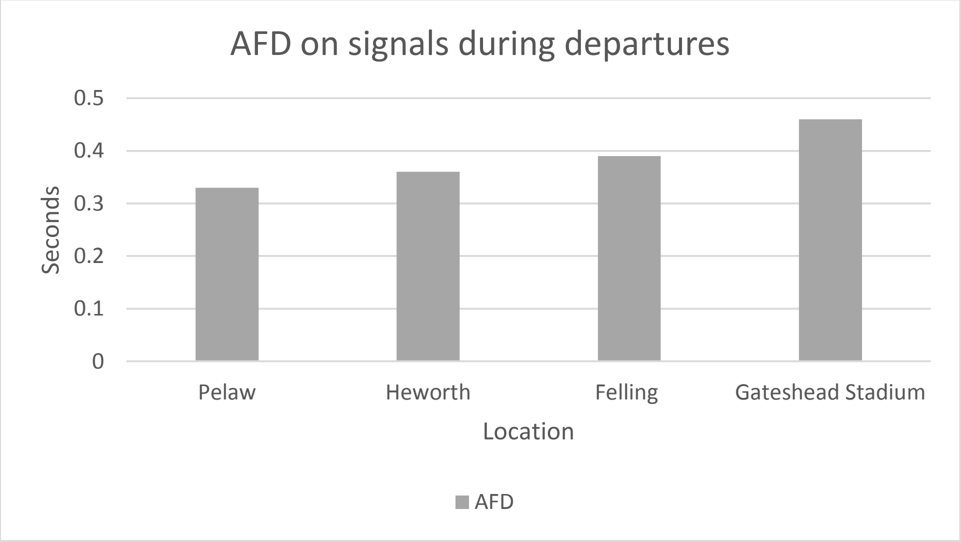


Figure 35. AFD on signals during departures

### 7.3.10 *Wilcoxon Signed Ranks Test for AOI on departure*

Table 23 below provides a summary of Wilcoxon Signed Ranks Test results for departure elements for all three metrics. Only one statistically significant difference is found between TFD on mirrors at Felling and Heworth when *p*-value < 0.05. Relationships for Pelaw-Felling signals TFD and Pelaw-Heworth signals TFC are close to statistical significance with p-values slightly higher than 0.05.

| | | Metric | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | TFD | | | TFC | | | AFD | | |
| | | Pelaw | Heworth | Felling | Pelaw | Heworth | Felling | Pelaw | Heworth | Felling |
| **Mirror** | Heworth | -1.475 | | | -.132 | | | -1.307 | | |
| | Felling | -.859 | -1.979** | | -1.518 | -1.507 | | -.075 | -.896 | |
| | G. Stad | -.709 | -1.307 | -1.083 | -.525 | -.787 | -.732 | -1.120 | -.411 | -1.568 |
| **Signal** | Heworth | -1.350 | | | -1.912* | | | -.631 | | |
| | Felling | -1.670* | -.187 | | -1.260 | -1.149 | | -.704 | -.280 | |
| | G. Stad | -1.006 | -.168 | -.161 | -1.082 | -1.391 | -.186 | -1.408 | -1.307 | -1.489 |

**Table 23. Difference in gaze behaviour at different locations (departures). *p<0.1, **p<0.05**

### 7.3.11 *Failure to check signals*

Analysing the descriptive statistics for the departure variables (histograms) as a part of normality tests revealed one major risk factors. The histograms for the signals show at least one occasion of not fixating at a signal at each of the stations.

In comparison to arrivals, Tyne & Wear Metro drivers are obliged to check a running signal twice before departing from a station – before closing doors and before pulling out of a station. The semi-interviews confirm that both of these checks usually happen within 10 seconds before the departure.

In total 11 occasions of no fixations at a signal are registered in 80 recorded departure routines. Pelaw has five incidents of this kind, Felling – three, Heworth – two and Gateshead Stadium – one. There is no individual offender with all the participants registering at least one violation. Taking into account significant experience of the participants, it is necessary to check effects of the sampling quality on these cases.

Even though it was mentioned before that the sampling measure was not found to have any effect on neither descriptive statistics nor statistical tests, these factors have influence on signal fixations on departure. 15% of trials (with <50% sampling measure) have 36% of failures to check a signal on departure. When such trials are excluded, the seven failures to check signals are divided this way: Pelaw – three, Felling and Heworth – two, Gateshead Stadium – zero.

### *7.3.12 Influence of personal and environmental factors*

The results of Kruskal Wallis H test for dependence on personal and environmental factors can be seen in Table 24 for TFD variables, Table 25 for TFC variables and Table 26 for AFD variables. The proportion of the variables found dependant on different factors is presented in Table 27 with respect to different metrics, AOIs and locations. Personal factors are found to have more effect on gaze behaviour than environmental factors, especially at arrivals. Dependence on environmental factors is more pronounced on departures. Expanding p-value to 0.1 increases number of significant relationships with personal factors more than with environmental. However, the number of new relationships uncovered is relatively small.

| TFD | | **Arrival** | | | | **Departure** | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Factor** | AOI | Pelaw | Heworth | Felling | G. Stad | Pelaw | Heworth | Felling | G. Stad |
| **Personal** | Mirror | 9.000** | 8.551** | 3.476 | 2.458 | 7.037* | 8.323* | 15.199** | 13.740** |
| | Platf. | 8.946** | 9.219* | 9.857** | 12.540** | | | | |
| | Signal | 15.506** | 1.78 | 5.175 | 6.271* | 2.244 | 3.900 | 2.619 | 4.633 |
| | StopP. | 10.269** | 13.632** | 11.939** | 11.690** | | | | |
| **Environ.** | Mirror | 1.709 | 2.161 | 2.304 | 3.071 | .654 | 4.789* | 5.859* | 2.578 |
| | Platf. | 4.477 | .732 | .058 | 3.369 | | | | |
| | Signal | .953 | 1.548 | .948 | 3.507 | 6.354** | 5.616* | 1.940 | 8.354** |
| | StopP. | 1.267 | 2.779 | 2.482 | 3.169 | | | | |

**Table 24. Dependence of TFD variables on personal and environmental factors. *p<0.1; **p<0.05**

| TFC | | Arrival | | | | Departure | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Factor** | AOI | Pelaw | Heworth | Felling | G. Stadium | Pelaw | Heworth | Felling | G. Stadium |
| **Personal** | Mirror | 3.665 | 8.708** | 1.279 | 6.268* | 5.305 | 7.638* | 12.629** | 4.650 |
| | Platf. | 10.426** | 10.177** | 6.396* | 8.591** | | | | |
| | Signal | 9.793** | .758 | 6.011 | 5.382 | 3.205 | 4.096 | 4.078 | 3.247 |
| | StopP. | 15.318** | 12.872** | 9.948** | 11.585** | | | | |
| **Environ.** | Mirror | 1.833 | 4.245 | .647 | .775 | 4.063 | .052 | 8.112** | 3.370 |
| | Platf. | 3.112 | .838 | .670 | 2.760 | | | | |
| | Signal | .352 | .962 | 1.275 | 3.189 | 1.037 | 6.044** | 1.671 | 3.153 |
| | StopP. | .834 | .982 | 8.205** | 1.304 | | | | |

**Table 25. Dependence of TFC variables on personal and environmental factors. *p<0.1; **p<0.05**

| AFD | | Arrival | | | | Departure | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Factor** | AOI | Pelaw | Heworth | Felling | G.Stadium | Pelaw | Heworth | Felling | G.Stadium |
| **Personal** | Mirror | 11.980** | 2.161 | 10.269** | 10.226** | 4.605 | 9.405** | 8.524** | 11.446** |
| | Platf. | 6.682* | 4.055 | 9.103** | 6.471* | | | | |
| | Signal | 9.654** | 1.790 | 4.083 | 6.442* | 1.887 | 3.156 | .621 | 1.853 |
| | StopP. | 5.002 | 10.318** | 5.049 | 5.009 | | | | |
| **Environ.** | Mirror | .826 | 1.011 | .857 | 3.149 | .401 | 1.929 | 6.354** | 1.083 |
| | Platf. | 4.557 | 1.149 | .144 | 1.521 | | | | |
| | Signal | .811 | 2.740 | .685 | 3.456 | 1.016 | 4.271 | 2.961 | 9.311** |
| | StopP. | .756 | 4.460 | .181 | 3.764 | | | | |

**Table 26. Dependence of AFD variables on personal and environmental factors. *p<0.1; **p<0.05**

| | **Personal factors** | | **Environmental factors** | |
|---|---|---|---|---|
| | Arrival | Departure | Arrival | Departure |
| **All metrics** | 54% | 29% | 2% | 25% |
| **TFD** | 69% | 38% | 0% | 25% |
| **TFC** | 56% | 13% | 4% | 25% |
| **AFD** | 44% | 38% | 0% | 25% |
| **Mirror** | 50% | 58% | 0% | 17% |
| **Platform** | 67% | | 0% | |
| **Signal** | 25% | 0% | 0% | 33% |
| **Stopping position** | 75% | | 8% | |
| **Pelaw** | 75% | 0% | 0% | 17% |
| **Heworth** | 58% | 33% | 0% | 17% |
| **Felling** | 42% | 50% | 8% | 33% |
| **Gateshead Stad.** | 42% | 33% | 0% | 33% |

**Table 27. Summary of importance of personal and environmental factors by metric, location and AOI ($p<0.05$ only)**

## 7.4 Discussion

### *7.4.1 Drivers' preferred areas of interest during arrivals*

The mirrors are the main area of visual attention (based on TFD) at all stations but Heworth. At Heworth the participants look at the platform for longer times than any other AOI. This is the station with the second highest passenger loadings among all selected locations. It is also the only station where the mirror is on the same side as the platform. Such layout means that the mirror can be obscured by passengers until later into an arrival timeframe. It is also possible that the two AOIs receive some error fixations due to being very close together. The overall statistics from other stations shows that the drivers use mirrors to determine a stopping point otherwise looking on those for so long is pointless on arrivals. It is necessary for them to select a stopping position that allows carrying out departure procedures safely, e.g. good visibility of the platform-train interface.

The drivers approximately spend 30% of their 15-seconds arrival sequence on either saccades or looking outside the AOIs. These outside areas could be a track, scenery or in-cab indicators. However, the more built up stations (Heworth and Gateshead Stadium) show higher TFD on all four AOIs, showing that the scenery fixations could account for up to two seconds at overground locations. Tyne & Wear Metro stations differ in terms of openness feel. The results of T-test confirm that there is a difference between open and confined layout stations. More condensed grouping of the AOIs influences the TFD due to reduction in saccadic movements and transition time between the design elements.

For open layout stations (Pelaw and Felling) a stopping position marker becomes the second most looked at AOI. According to GI/RT7033 (Railway Group Standards, 2009) the recommended dimensions for platform stop markers is only 300x250mm with visibility requirement of only two seconds. The results show the metro drivers fixating on all of the markers for more than two seconds and those being visible significantly earlier. This shows an important difference between conventional mainline railways and DOO systems where it is necessary to provide drivers with increased flexibility of stopping positions. Moreover, when the DOO equipment is used as a stopping position benchmark, 2-second visibility is too small as it does not factor issues of divided attention between the mirrors and the markers.

In absolute values, the TFD for a stopping position indicator is slightly lower at Gateshead Stadium even though the station also has an island platform. The confined nature of the station contributes to more gazes at the platform. It was observed during the tests that the drivers do not use the marker at Gateshead Stadium as a stopping position. They usually stop further down the platform, with the first set of doors being next to the marker. It was not the case at other locations. With the participants prioritising the mirrors as a stopping aide, the incorrectly positioned stopping marker becomes only a general benchmark and loses its informativeness. Total fixation duration tends to decrease with loss of informativeness (Afrooz *et al.*, 2014). Drivers’ attention to the platform at Gateshead Stadium can be caused by a passenger access ramp on the right hand side and a retaining wall on a left hand side. These design elements along with a bridge in front and a ticketing area on top create confined physical environment where the same trends as at Heworth apply.

The participants fixate the least on running signals which is unexpected as the drivers clearly indicated checking the signals on arrival in the questionnaires (Section 6.3.1). This discrepancy indicates that the drivers might have been trying to answer the questionnaire in a way they thought to be correct. Morrel-Samuels (2002) claims that such behaviour is not unique and answer favourable skewness can be a sign of anonymity concerns. However, in a more demanding environment of an operational railway, checking something which is not required can be skipped in order to address more critical tasks.

The lowest signal TFD is observed at Gateshead stadium. The running signal there is located on the longitudinal point as the mirror. As the drivers get closer to a stopping point, the angle between the signal, the cab and the mirror becomes wider. The concept of such angle is demonstrated in Figure 36, where it is depicted in red. As the angle gets wider, the signal and the mirror move further apart in a driver's field of view (FOW). Hence saccadic movements become longer and less time can be devoted to fixations. The FOW angle is smaller at Felling than any other station with an island platform and it shows in the highest TFD on a running signal.

The smallest FOW angle can be seen at Heworth where a driver's cab is on the same side as the mirror and the signal, but it does not result in the highest TFD. It is possible that the total distance of the signal from the stopping position is simply too big at Heworth making this AOI less relevant for the drivers during the stopping procedure. Groeger *et al.* (2001) showed that mainline train drivers fixate on a signal 8 seconds before reaching it, which means at least 250 m (at 100 km/h). On entering Heworth station (from a curve) drivers are never more than 150 m away from a signal. However, the research in mainline railways only considers between station signals which do not have to compete with the PTI-related tasks. Hence it is possible to assume that signal sighting distance on station approaches, especially in DOO systems, reduces significantly.

It is also important to note that due to relatively low values for TFD on running signals, these results should be taken with caution. The eye-tracker's low sampling frequency can cause an error of up to 33ms which is more than 15% in case of Gateshead Stadium signal. The results are not supported by statistical tests too. It was also expected for the drivers to fixate more frequently on Pelaw signal due to convergence of 3 lines (including sidings) right before the station thus higher informativeness of this signal. However, the results report that drivers treat all of the signals similarly in terms of SA building.

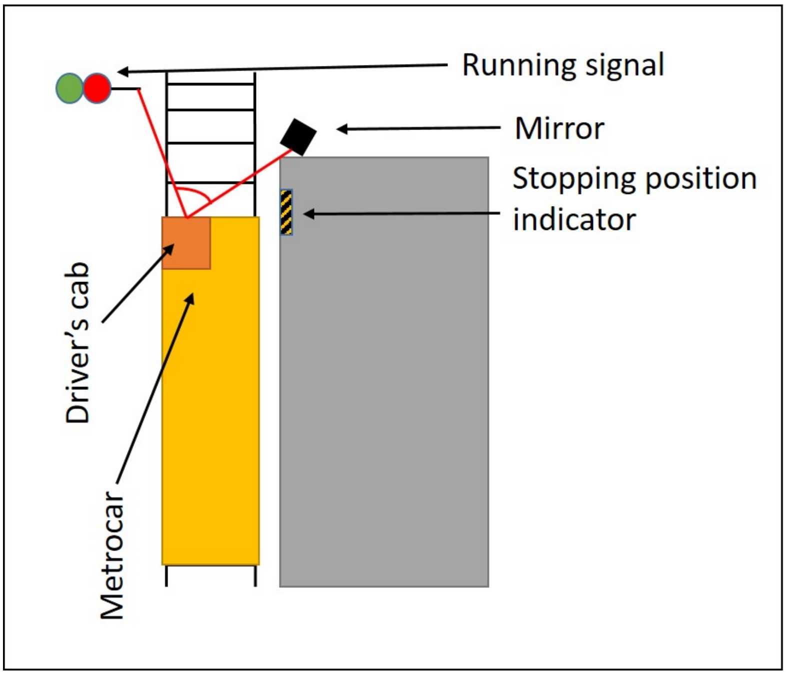


**Figure 36. FOW Angle between a signal, a driver and a mirror (in red)**

### *7.4.2 Elements inducing additional stress and workload during arrivals*

The methodology to test for stress-inducing physical environment is largely undeveloped in human factors domain. The high TFC/low AFD method used in this study is based on a collection of previous work which sometimes suggested controversial results. However, the author believes that this method is sufficient for exploratory study where only relative changes in usability are considered. This research area will definitely benefit from large scale testing which aims at establishing baseline values for eye-tracking metrics in usability studies. Moreover, agreeing the baseline values for "normal" operations will open endless possibilities for assessments of drivers using the eye-tracking devices (Milleville-Pennel *et al.*, 2010).

In terms of mirrors, the open stations show signs of poorer usability. As other two stations in the study have the mirrors either under a structure or protected from direct sunlight or side wind, it is possible to assume that environmental factors are rather important. Even though the environmental factor assessment does not support these assumptions, there are fundamental flaws with the assessment as described below. The second PSF differentiating between the mirrors at open and confined stations is the slightly higher distance to the mirror at the former locations. With the mirrors of the same size, it can be expected for shorter distances to provide better views. On the other hand, it is unknown why the drivers do not ignore pre-defined stopping locations at the open stations similarly to Gateshead Stadium.

As drivers constantly switch their attention between the stopping position markers and the mirrors on arrival, usability of one should affect another. However, this relationship might not be as straightforward or proportional as one can expect. For example, the stopping indicator with poor informativeness at Gateshead Stadium allows drivers to interact easier with the mirror. On the other hand, the lower TFD at Heworth mirror (as the result of more attention attribution to the platform) means relatively better usability of the stopping position indicator. The stations, where TFD is high (compared to other stations) on both mirrors and the indicators, suffer from reduced usability of the main stopping-related element – the mirror. These observations suggest that low informativeness (and subsequent lower TFD) of one of the 2 AOIs responsible for safe stopping actually results in drivers prioritising their attention on only one of the elements. This helps avoiding a divided attention issue which is known to significantly increase stress and workload (Davies *et al.*, 2013; Prottengeier *et al.*, 2016). It is important to note that station arrival encompasses more than 2 concurrent monitoring tasks. Metro drivers also need to observe the PTI, control train's motion, monitor the track for objects. Furthermore, stopping a train is also associated with increase in emotional workload (Myrtek *et al.*, 1994). Even though Basacik *et al.* (2015) claim that station stop associated tasks can cause underload, neither of their "underload" scenarios is applicable to DOO systems.

In terms of platforms, only Pelaw demonstrates signs of poor usability. These are based on only a marginal difference in AFD with Felling. The literature does not specify whether it is necessary to compare TFC/AFD ratios in situations when one of the metrics produces the same values. However, as discussed in the methodology (Section 7.2.4), both metrics can report the same in certain scenarios. For example, high TFC can be a sign of difficulty in information extraction (Ehmke and Wilson, 2007; Holmqvist *et al.*, 2011) but is more controversial when not supported by low AFD value. One factor which might have affected the results at Pelaw is location of the waiting area and consequent higher concentration of the passengers towards the back of the train. Concertation of passengers in only one part of a platform is a known risk which can be caused by station design (Cynk *et al.*, 2015). At Pelaw, this leads to the drivers needing to scan large volumes of passengers on higher speeds. Fitzpatrick *et al.* (2010) showed that multi-tasking on higher speeds causes decline in car drivers' performance. However, this is only applicable to the situation at Pelaw if observing each individual passenger is considered as a separate task.

The relationship between the signals at Heworth and Felling needs to be considered with caution due to factors described above. If the signal at Felling is in fact easier to interact with than Heworth signal, the distance between a stopping position and a signal becomes important PSF. However, the adverse effects are not directly proportional to the distance and only become pronounced after certain distance. Based on other locations, it is possible to note that distances in excess of 33m (Felling) are where this process starts.

### *7.4.3 Trends in fixations during departures*

The statistical tests confirm the highest TFD on the AOIs at Heworth and the lowest TFD on the AOIs at Felling. The tests also show Gateshead Stadium having the second highest TFD on the AOIs. Once again, more confined stations get the highest gaze durations. For these stations increased TFD comes from longer fixations on the mirrors. Both Heworth and Gateshead Stadium have shorter distances between the mirrors and the stopping points.

The drivers fixate on the mirrors for significantly longer than on the signals. It is possible that the drivers assume signals as a static element. If it is green (checked on approach or before the last 10 seconds) it normally will stay green. The area under investigation has no converging lines and turnback facilities only at Heworth. The mirrors represent dynamic environment which requires significantly more monitoring.

Visual attention allocated to the AOIs is strongly associated with the FOW angle. The wide angle layouts primarily cause a drop in TFD on signals, which is shown by Gateshead Stadium and Pelaw. It is clear that drivers are willing to sacrifice signal checks to some extent in layouts requiring longer saccadic movements. In comparison to the indicator-mirror relationship on arrivals, the two AOIs on departures are both safety-critical and, in theory, cannot be skipped. Hence the issue of divided attention is further prompted by certain layouts and PSFs. This can manifest itself in more stressful interaction with one of the AOIs, as demonstrated by Pelaw mirror.

### *7.4.4 Elements inducing additional stress and workload during departures*

The mirror at Pelaw has been highlighted as a problematic area before due to potential environmental factors. Similar layouts are used at Felling and Gateshead Stadium albeit a narrower FOW angle, and do not demonstrate mirror usability issues. This corroborates importance of the FOW angle as PSF. Other factors, e.g. crowding at the back of the train and a closely located mirror can contribute to workload increase too. However, it is unknown to what extent each of these factors influence the usability of the mirror.

Similarly to the arrivals, results for the running signal at Heworth are indicative of stress-inducing PSFs. Compared to other stations, it is located the furthest from the drivers. Assuming that the low TFC/high AFD suggests better usability of an element, there is a negative correlation between the distance to a running signal and signal's usability.

### *7.4.5 Failures to check signals during departures*

As it was mentioned before, start against a signal SPaDs are rather common in railway industry, especially on DOO commuter routes, and Tyne & Wear Metro is no exception. According to RSSB (2013), there were at least 39 SASSPaDs with passenger trains in the UK between 2010 and 2013. Moreover, SASSPaD risk is higher in DOO systems according to Basacik *et al.* (2009). As presented in Precursor Incident Model by RSSB (2011b), many of driver-related accidents stem from smaller incident acting as precursors. In this particular situation, failures to check a signal aspect before departure can lead to SASSPaD under certain circumstances. However, it is necessary to assess how real is such proportion of violations on departure.

The percentage of the violations found in trials with less than 50% sampling measure is out of proportion. All of the registered violations in those trials are recorded at stations with an island platform. Hence it is possible that the drivers were using side gazes to check a signal due to mainly focusing on a mirror. Such side gazes are rarely recorded, especially if an eye-tracking set is a little bit disturbed on a participant's head. The same issue was observed by RSSB (2005) who reported a lot of lost data due to glares, equipment failures and side gazes. Drivers' main areas of attention are located on the same side on arrivals. Hence no difference in descriptive statistics has been found when the 3 trials with low sampling quality are excluded from the analysis. This could be a reason for low TFC on arrival signals. However, this theory would mean Heworth having significantly higher signal metrics on arrivals, which is not the case.

Apart from technical issues embedded in methodology, it is known for people visual attention to travel ahead of fixations by about 0.25 seconds (Deubel, 2008; Holmqvist *et al.*, 2011). Hence if participants wanted to fixate on a signal right before departure, the actual fixation could have happened after a train started moving (not analysed). Considering that there is a slight lag between applying a power and start of motion, it is possible to claim that the lag between visual attention and fixations is negligible for this experiment.

Most importantly, whatever the reason for equipment not registering a fixation on a signal, even if one of these cases is a genuine violation or a lapse, this creates a significant risk factor for metro operations. This is especially important when lack of fixations on a running signal on arrival is taken into account. Even checking a signal 10 seconds before departing, if other tasks are done in-between, cannot be safe. O'Connell *et al.* (2015) claim that dynamic occurrence of task demands can impact on individual's ability to recall previous observation.

Furthermore, the participants do not differ in ways how they interacted with the signals, especially on departure, meaning that if there is a problem, it is wide-spread.

After the questionable violations are removed, the seven violations left are at stations where a running signal is located at least 30 metres away from a stopping position. It takes 7 to 10 seconds to cover 32m and 45m respectively metres from a complete stop. Hence the drivers, especially experienced drivers, might believe that they have time to check a signal and react even if they do it only after departure. However, Multer *et al.* (2015) claim that very responsive controls of electric trains mean more in-cab gazes for speed information monitoring and subsequent higher risk of SPaDs. If the drivers decide to divide the two tasks (safe dispatch and signal checking), they might find themselves in situations when other driving tasks draw their attention when a train is already in motion. It is also possible that the station stops and concurrent PTI demands simply cause a memory lapse, as station stops can cause disengagement from a driving task (Naweed and Rainbird, 2014).

RSSB (2005) found that drivers' experience does not affect visual behaviour on signals but only between station signals were checked. Tripathi and Borrion (2016) discovered that emphasising on-time performance affects metro drivers' safety and security related performance. It also means that other competing goals can have similar effect on metro drivers' performance. RSSB (2007b) claim that drivers are highly motivated to check signals on departure and non-compliance is highly unlikely. On the other hand, another RSSB report highlights a risk of drivers departing a station without fully checking all doors closed due to time pressures (Basacik *et al.*, 2009). In DOO systems, where most risks are related to the PTI, the drivers might prioritise DOO equipment and skip signals due to other pressures. This demonstrates why same findings cannot always be applied to mainline railways and metro systems as a combination of design features and risk profiles can significantly affect drivers' priorities. This is further supported by five out seven violations happening at the stations with significantly higher passenger loadings.

As bigger distances do not correlate with increased number of failures to check a signal, it is possible that another design-related factor can have influence on failure propagation. This factor is the FOW angle. Felling and Pelaw, which have relatively high angles and distances, show the highest proportion of violations. In comparison, Heworth, where the distance is also high but the angle is significantly lower, does not have as many failures to check a signal.

### *7.4.6 Personal and environmental factors*

The personal factors show considerably higher effect on drivers' gaze performance compared to environmental factors, which are almost negligible. Each individual participant has different strategy looking on the most dynamic AOIs – platforms on arrival and mirrors on departure. Varied passenger levels might have contributed to this. This is supported by the fact that during arrivals the variability is mostly observed at the two most crowded stations – Pelaw and Heworth. The statistics also shows that number of passengers on platforms can influence interaction with other elements, namely stopping position indicators and potentially mirrors. The participants who are more cautious about the platform crowding might spend more time (TFD most varied metric) looking on platforms causing TFD on other AOIs to reduce. The results corroborate with previous research emphasising high importance of personal factors (Li, 2004; Gibson, 2016).

AFD and TFC metrics demonstrate significantly lower percentage of statistically significant relationships with the personal factors than the TFD values. This is a positive result for the methodology taking into account concerns about assessment methodology of stress-inducing elements. However, it is important to note that different drivers did different number of trial runs which could have affected some of the results. Moreover, passenger levels encountered could be very different as this not only depends on time of a trial but also presence of peak services (relieving passenger congestion) ahead and disruptions in the system.

The environmental factors' influence was rather low implying that sun is not the contributory factor at Pelaw for the mirror during departures. However, the drivers' in their questionnaire assessment of contributory factors stated that the direct sunlight is the issue, not sun in general. As the trials were conducted in different time and different months, it is impossible to say in retrospective whether some of the trials had that issue or not. On the other hand, a measure for wind was not recorded. The issue of "shaky" mirrors was raised during the semi-structured interviews by the drivers. As it can be very windy in Tyne & Wear during any type of weather, the wind should be still considered a major contributory factor to poor usability of the mirrors at the open stations.

#### *7.4.7 Lack of statistically significant differences in mean values*

Some of the relationships discussed above are not statistically significant, i.e. for signals and stopping position indicators. The results were double-checked using Sign Test, which returned the same statistically significant relationships. Firstly, it is possible that the sample size was simply not big enough, which is known to have effect on statistical significance (Section 7.2.7). However, the expansion of significance parameter allowed unveiling more relationships where difference in variables is pronounced. Secondly, the selected eye-tracking device potentially created a lot of deviation in the results making harder to find statistical significance. This could have been caused by the necessity to explore dynamic physical environment using static AOIs, low sampling rate and lower sampling measures in some trials.

The analysis supported some pronounced differences in descriptive statistics reinforcing assumptions. If differences with $p<0.1$ are taken into account, 25% of arrival relationships would be found statistically significant. It cannot be expected that all of the AOIs show difference in all metrics. Moreover, the exploratory nature of this experiment showed the areas that need to be explored even if those were not highlighted as statistically significant at all locations. However, lack of simultaneous statistically significant differences in both AFD and TFC variables means that all assumptions about increased workload/stress are purely based on descriptive statistics.

## 7.5 Conclusions

The research community does not use eye-tracking to the same extent as other safety-critical industries leaving a serious gap in the body of knowledge. The study has shown that the eye-tracking research can be conducted safely in metro systems. It allows building a basis of the exploratory research in front-line staff interaction with the physical environment. Moreover, the more sophisticated eye-tracking tools available now and advancements in experiment methodologies should allow establishing baseline eye fixation values in future. Such baseline values would facilitate assessing the infrastructure and human operators in a non-intrusive way in order to explore performance.

Metro drivers prioritise mirrors as the main AOI for visual attention. This suggests bigger semantic value of the PTI risks than SPAD risks to them. Arrivals see the drivers' visual attention divided between the mirrors and the stopping position indicators, but gaze duration on platforms depends on passenger levels and openness of a station. When one of the elements in this relationship exhibits poor informativeness, the issue of divided attention is not as acute as the drivers focus on the associated AOI.

In terms of workload and stress inducing PSFs, several design features are proposed by the data. Firstly, a wide angle between a mirror, a driver and a signal (FOW Angle) creates situations when drivers need to produce longer saccadic movements to shift their attention from one AOI to another. Secondly, the distance between a signal and a stopping position affects drivers' interaction with the signals. Thirdly, the passenger levels and distribution on a platform can create additional stress for drivers during both arrivals and departures. However, the analysis shows that different drivers approach complications caused by crowding differently. Albeit almost negligible influence of weather (sunny or rainy) on drivers' performance, other environmental factors, e.g. wind or direct sunlight, can still affect drivers' visual interaction with the physical environment. This is demonstrated by differences between open and built over stations as well as drivers' perceptions.

The biggest concerns are related to drivers potentially not checking a signal before powering up the train. This is a serious precursor in SaSSPaDs and should be addressed. Even though the data could have been affected by quality issues, the absolute number is too high to be disregarded. In fact, even one occasion in 80 station departures is alarming as all of the participants are very experienced drivers.

With many tasks competing for metro drivers’ visual attention in DOO systems, the drivers seem to prioritise monitoring of the PTI over mitigation of SPaD risk. This is significantly different from conventional mainline railways, where a train guard or platform staff ensure safe dispatch of a train.

Combination of the distance to a signal, the FOW angle, and passenger loadings seem to have effect on propagation of such violations. The drivers have up to 10 seconds between powering up and passing a signal, which can create a perception that there is a room to mitigate an error. Moreover, there is a possibility that such perception is developed by more experienced staff who understand better such technicalities of the system.

Due to exploratory character of this experiment, the next step is to confirm the influence of the discussed PSFs on metro drivers’ performance. The investigation should be carried out in a more controlled environment where several design features can be tested against each other. This is hard to accomplish in another in-field study as it would involve altering infrastructure of a running railway with all the associated risks. However, a simulator study is a useful approach to achieve this. The simulation must include a more uniform sample with participants doing comparable number of trials in order to avoid uncertainties placed by personal factors on descriptive statistics.

# Chapter 8. Simulator study

## 8.1 Introduction

With the conclusions of the in-field experiment in mind (Chapter 7), a simulator study has been conducted to validate the findings discussed in the previous chapter. Simulators have been evaluated in context of railway industry by various national research bodies before (Rushby and Seabrook, 2007; Jolly *et al.*, 2013). Presenting a case for simulation, Rushby and Seabrook (2007) list several advantages of this method, including cost and operational effectiveness, ethicality of assessment and repeatability of environment. In many studies simulator use is the only way how to re-create certain scenarios without serious disruptions to service. Even though the evaluation is focused on training needs, simulator features required for training coincide with simulator features required for assessment of performance shaping factors. As with eye-tracking and general performance shaping factors research, the rail industry is somewhat behind automotive industry in terms of simulator use. This is why the next paragraphs have many references of research in car driving. However, the use of simulators in railway-related research is gaining momentum and becomes more wide-spread.

According to Jamson and Jamson (2010) each simulator can be assessed on five parameters: motivation, physical validity, face validity (visuals and controls), perception (auditory, ocular, proprioceptive factors helping to judge distances, speeds) and behavioural (perception of environment). Kaiser and Schroeder (2003) provide slightly different model of four parameters: physical, visual, motion and cognitive fidelities. According to Jolly *et al.* (2013), the rail simulators can be divided into several categories: full simulation (full cab), part-task trainer, PC-based and role play. In automotive industry the simulators are usually divided into low, medium and high level. This categorisation is based on simulator's similarity to real-life in comparison to function-driven categorisation in the rail industry. The closer the simulator is to low level fidelity, the more it compromises on different parameters, e.g. face validity or physical characteristics. It is important to note that the literature review did not find any certain guidelines on what constitutes low or high level simulator. General assumption is that a high level will be a full car simulator with a moving base, medium – a full car on a fixed base, whereas a low level simulator will not have immediate physical environment re-created.

High cost of high fidelity simulators prompted a lot of research into cheaper solutions and validity of those. Yates and Sharples (2009) showed that a simulator consisting of a laptop and a wide field of view projector produced more realistic results than a full-cab trial. In the automotive industry, Jamson and Jamson (2010) and Reed and Green (2010)  showed that a simulator built using just a PC (17" screen) with a gaming steering wheel reproduced drivers' speed performance in the same way as a high level simulator. However, same was not achieved for lateral performance. One of the main differences between the low and high level simulators is usually lack of wide field of view in former, which can negatively affect drivers' performance (Kappé *et al.*, 1999). In regards to fixed base simulators, Bella (2014) claims that such simulators have proven their reliability in many studies before. It is especially applicable to studies which are interested in relative validity Törnros (1998). Lemieux *et al.* (2014) demonstrated that car drivers' perception of the simulator and subsequent performance does not differ between mid- and low-level simulators. Knapper *et al.* (2015), in their comprehensive assessment of the previous research in simulator validity, showed that both high and low level car driving simulators can show both good and poor absolute and relative validity. According to Jamson and Motula (2004), less expensive analogues can be successfully used for certain tasks without difference in participants' response. Interestingly, Jolly *et al.* (2013) claim that the most important benefit of high fidelity simulators is ability to put users under realistic workloads stresses.

Even with high level simulators, there are concerns of lack of motivational pressures in extreme situations (Knapper *et al.*, 2015; Tripathi and Borrion, 2016). The research from other industries seem to demonstrate no difference in workload between real life and simulated environments but concerns are expressed on how to check whether this similarity still holds in emergency events (Jolly *et al.*, 2013). Moreover, many simulators focus on only a certain drivers' task thus jeopardising the ability to generalise the results. According to Knapper *et al.* (2015), driving the simulator also involves a steep learning curve which can affect the results. Moreover, not many drivers are used to drive while being observed thus they might start adjusting their behaviour to what they believe is desirable by an observer.

There are still concerns about validity of results of simulator studies in real world but it does not stop the simulators from being used in drivers' performance studies. In the UK simulation studies were conducted on drivers' interaction with head-up displays

(Davies *et al.*, 2007), in-cab signalling (Buksh *et al.*, 2013), effects of risk triggered commentaries (Sato and Bowler, 2015), differences in performance between train drivers and quickly taught non-driver (Large *et al.*, 2015). Basacik *et al.* (2015) studied what train driving situational factors lead to underload and undemanding working conditions. Aitchinson and Davies (2009) used a high fidelity train simulator to study drivers' performance with new junction signalling layouts and designs.

Outside of the UK, Tripathi and Borrion (2016) studied potential trade-offs made by metro drivers who are pressured for on-time performance in a high fidelity metro simulator. Train drivers' fatigue was investigated in simulator studies by Dorrian *et al.* (2007). Verstappen (2015) used medium level simulator to study how a presence of a second person in a cab affects train drivers' performance. Naweed and Balakrishnan (2013) looked into different strands of fidelity in an effort to evaluate rail simulator fidelities.

Albeit rarely but the simulation studies have been used to evaluate physical environment, namely infrastructure. De Ceunynck *et al.* (2015) advocates that simulators can have an important role in pro-active evaluation of drivers' physical environment and its usability. Their research allowed introducing more than 20 improvements to road signs and markings used in a highway reconstruction project. Van Luipen and Meijer (2009) used their own developed gaming simulator to study feasibility of infrastructure alterations and interventions into railway physical environment. Rentzsch *et al.* (2009) asked drivers from different EU countries to drive a train simulator in order to explore design of a universal pan-european train driver cab design.

The studies described above used a mix of subjective and objective methods to assess drivers' performance in a simulated environment. Eye-tracking is one of such non-intrusive methods. Many researchers collect speed and other measures from the simulator. For many a simulation study is only a part of a bigger research project thus it is approached as a hypotheses testing experiment. However, most studies do not include any work into validation of a simulator even when questions are raised about results validity. A cross-industrial literature review is usually used as only means of methodology validation.

With the described complexities in simulator development and costs associated with it, it is no surprise that there is a drive for more cost-effective simulators. It is not only about substituting high level setups with low or medium level installations. Santiago-Chaparro *et al.* (2011) outlined how widely available datasets and software packages can be used in order to improve efficiency of simulation scenario creations. Even though a general assumption of a rail simulator is a high to medium level simulator, Rushby and Seabrook (2007) claim that one of the available PC-based gaming train simulators (at that time) had potential to be used as part task trainer. For example, Rowe (2013) used a gaming engine to build a route learning trainer for Manchester Tramlink.

## 8.2 Simulator

### *8.2.1 The virtual reality*

Swift progress of gaming technologies and open access availability of such gaming engines as Unity or Unreal Engine 4, and graphic design tools as Blender allow almost everyone to design high detail virtual realities with a limited cost. It was decided to explore research potential of ready-made gaming software, i.e. Train Simulator 2016 from Kuju Entertainment. The choice of software is based on several reasons. Firstly, the game provides a variety of very realistic rolling stock which have good physical and face fidelities, fully re-creating train cabs. This can substitute building a mock-up of a cab as the virtual reality already has one. Secondly, a built-in scenario editor allows developing virtual railway systems quickly. The asset base available is large and continues to grow due to a big game fan community. Thirdly, the game allows importing own created objects from easy-to-use graphic design packages, e.g. Google Sketch Up. With these advantages in mind, the part of the system featuring in the in-field experiment (Pelaw to Gateshead Stadium) was built in virtual reality in several steps:

- Tracks laid using Google Maps satellite images,
- Models of the stations were designed in Google Sketch Up or Blender in 1:1 scale,
- The models were textured using real-life photos of the stations in 3dMax 2015 or Blender,
- The models imported into the game and positioned using Google Maps satellite images for correct location,
- In-game assets were used for track-side equipment, DOO mirrors, scenery and passengers. The locations of assets were determined using site plans and real-life cab front-facing cameras’ recordings.

Creating 3 km long route using this method can take between one day and one month depending on previous experience and availability of drawings, textures and other support materials. The use of real-life textures helps avoiding problems mentioned by Thomas and Gassel (2009) who claimed that simulators provide realistic movements but retain impression of computer-generated environment. The simulated scenarios should be filled with real-world objects and scenarios to improve the realism (Bella, 2005; Yan *et al.*, 2008; Santiago-Chaparro *et al.*, 2011). Hence even posters at virtual stations were from real-life advertisement campaigns.

Screenshots of the virtual Tyne & Wear Metro can be found in Appendix J.

### *8.2.2 Simulator venue and setup*

The decision theatre at Science City in Newcastle upon Tyne was used to conduct the study. The facility has a six metres wide screen using two projectors with combined resolution of 3208x1200 pixels at 120 Hz, but the picture was upscaled from 1920x1080 pixels resolution for better frame rate. The picture is blended with a custom software tool to achieve seamless overlap. Participants were seated approximately three metres from centre of the screen. The room arrangements can be seen in Figure 37. It is important to note that in this picture the simulator is not in the full-screen mode for testing purposes.

**Figure 37. Author's colleague trialling the simulator setup. The head tracking camera is located on the chair under the screen.**

One of the main advantages of the high fidelity simulators is immersive surround screens which can accommodate for drivers' head movements. As an alternative to that, it was decided to pair the available screen with a head-tracking device supported by the software. The device (TrackIR 5 Pro) consists of an infra-red (IR) camera fixed under the screen and tracking sensors fitted either on a baseball cap or a headset. According to the developer website (NaturalPoint Inc, 2016) the device tracks motion in six degrees of freedom. This includes all three axes, yaw, pitch, and roll. Using TrackIR 5 Pro allowed realistically recreating situations where drivers have to lean sideways in order to see a mirror or a stopping position indicator. Moreover, there are several other significant benefits. Firstly, the immediate physical environment can be re-created in virtual reality thus saving money and increasing testing flexibility. Secondly, it minimises effects of so-called “simulator sickness”. This sickness is caused by a stationary user being exposed to a sense of motion (LaViola, 2000). Additionally, the IR camera mounted on a chair right below the screen also acts as a stationary “anchor” object to reduce the effects of the simulator sickness even further (Duh *et al.*, 2004).

As the virtual reality provides cab environment, which can be observed by a driver, it was decided to use a game console controller (Microsoft Xbox 360). Such control method has been selected to allow drivers maintaining visual focus on the physical environment instead of looking at the controller, e.g. keyboard. The controller has trigger buttons on the back which are used to interact with a master controller of a train. The controller also allows operating doors, window vipers, horns, AWS even though neither of these actions were required from the participants during the tests. Use of non-standard desk controls presented the biggest uncertainty in the study. Yates and Sharples (2009) report that there is evidence that non-realistic train controls could detract participants from performance. However, PC-based simulator by Rowe (2013) used keyboard to control tram movements and received positive assessment from participants for tram’s controls. As the literature review in Section 8.1 demonstrates, low level simulators compromising on immediate physical environment still show good relative, and sometimes absolute, validity. On the other hand, those studies are mostly focused on car driving where even a low level simulation can have a cheap gaming steering wheel.

### *8.2.3 Limitations*

The selected simulator methodology has several limitations. Firstly, Train Simulator 2016 does not have Tyne & Wear Metro specific rolling stock so alternatives had to be assessed. It was decided against designing design a custom-made cab as this would inflate scope. Instead, The Class 170 “TurboStar” train was selected as the rolling stock for the simulator. It has a driver’s seat on the left hand side, similarly to a metrocar, and with a right hand side blind partially lowered (Figure 38) it provides similar visibility through the windscreen. The “TurboStar” also has a master controller similar to a metrocar. However, the physics of the selected rolling stock (weight, acceleration, braking) were altered in the way to match metrocars’ performance. This performance was tested with two Tyne & Wear operations managers who confirmed that it is very realistic.

Figure 38. View of Heworth station from the cab. Right-hand side windscreen is partially obscured by a lowered blind.

Secondly, the gaming engine used by Train Simulator does not offer reflective DOO mirrors. Even though there are models of DOO mirrors in the software, those do not provide any information on PTI. One can only assume it is done to improve framerate. The in-field eye-tracking experiments demonstrated that the DOO mirrors contain the highest informativeness and semantic value for the drivers during arrivals and departures. Hence those are being constantly monitored. Jolly *et al.* (2013) showed that as many activities as possible need to be addressed in any level of simulation to re-create drivers' experience as much as possible. In order to maintain the same station duties workload and motivational pressures, the DOO mirrors were fitted with analogue clock faces of a size that is definitely distinguishable from a stopping position (Figure 39). The drivers were then asked to choose a stopping position where they clearly can read the time and call it out loud. For departures, the drivers were required to wait until a half or a full minute to depart.

Figure 39. An example of a DOO mirror fitted with an analogue clock

## 8.3 Methodology

### *8.3.1 Hypotheses*

With the in-field eye-tracking study in mind, the simulation study was designed to test five hypotheses. Namely:

8a. A shorter distance between a stopping position and a running signal leads to drivers fixating on the signal for longer and induces less stress/workload.

8b. A narrower angle between a mirror, a driver's cab and a signal (FOW Angle) positively affects ease of interaction with the physical environment during arrivals and departures.

8c. The combination of a lower distance to a running signal and a narrower FOW angle leads to fewer failures when checking a signal aspect during departures.

8d. Increased passenger loadings negatively affect workload when interacting with the areas of interest not related to platforms.

8e. Novice drivers (less than two years of experience) have different gaze behaviour compared to more experienced colleagues and are less prone to missing a signal check on departure.

### *8.3.2 Station design and alterations*

To test the hypotheses 8a to 8e in a controlled environment, three scenarios were developed re-creating the same part of Tyne & Wear Metro – Pelaw to Gateshead Stadium. The versions only differ in one element at each station in order to focus only on effects of one change at a time. The scenarios are:

1. **Baseline system**. This VR metro system re-creates the real-life counterpart as close as possible and was assumed to be a base level. The layout of the stations can be seen in Section 7.2.1.
2. **Amended system**. Baseline system with one of the elements of the physical environment amended at each station. See Table 28 for the list of amendments.
3. **System with higher passenger levels (more PAX)**. Baseline system with increased passenger levels, which represent loadings closer to real world peak levels. Appendices J and K show how the stations differ between the scenarios 1 and 3.

| Station | Element | Change from the baseline |
|---|---|---|
| **Pelaw (Type 2)** | Mirror | The element is moved 5 m further away from the usual train stop position (stopping position marker stays in the same location) |
| **Heworth (Type 1)** | Signal | The running signal is moved 25m closer to the stopping point |
| **Felling (Type 2)** | Signal | The running signal is moved 18m closer to the stopping point |
| **Gateshead Stadium (Type 2)** | Signal | The running signal is moved 25m further away from the stopping point |

**Table 28 List of introduced changes to station layout in scenario 2**

The scenarios were presented to the participating metro drivers in random order to avoid learning effect in multi-trial experiment (Yates and Sharples, 2009). The experiment lasted for approximately one hour with short brakes between each test in order to load the next scenario and diagnose the eye-tracker if necessary. The drivers are used to short driving portions as some peak services duties involve only 25 to 40 minutes long journeys with up to 10 minutes turn-around brakes.

### *8.3.3 Participants*

Five drivers took part in the simulation study making a total of 38 runs on the simulated part of Tyne & Wear Metro. Most of the studies identified in the literature review had less than 10 participants but none actually asked the participants to do more than one test. On the other hand, each participant did approximately the same number of tests thus creating a more uniform sample than the sample in the in-field study. Two participants were experienced driver coaches who also participated in the real-life eye-tracking study. Other three drivers had been driving the metro trains for less than two years at the time of the study and had been recruited for the tests on the basis of their inexperience.

### *8.3.4 Data sample*

It was envisaged that each driver would do an equal number of trials in each scenario. This would allow creating a matched sample where drivers' trials under each scenario are matched against each other. In total 42 trial runs were conducted with 5 participants creating a sample of 14 tests (three scenarios in each). Each of the trial runs consists of 4 successful station stops and departures.

### *8.3.5 Data collection and analysis*

The gaze data has been collected using the same eye-tracking methodology as in the in-field experiment (Section 7.2.5). The eye-tracking techniques were successfully combined with simulator studies before (Dutta *et al.*, 2004). There were concerns about interference between the IR head-tracking camera and the eye-tracking glasses but those devices do not affect each other. It is important to note a number of differences between the in-field and simulation studies, which might have implications:

- There is a possibility that the alterations in scenario 2 could introduce changes to criticality of AOIs for the drivers. Hence higher TFC/higher AFD situations should be considered from higher TFD point of view indicating lack of situational awareness with this layout (Ottati *et al.*, 1999).
- Better expertise (higher TFC/lower AFD) might become a factor in this study as there is a mix of experienced and in-experienced drivers. However, testing for experience effects of the eye-tracking measures should provide a guideline on how to assess such cases.
- Similarly to the previous point, the experience is not fully controlled in this experiment as scenario 2 is novel to both groups of drivers. Hence dwell times and fixation counts can be affected by this factor too.

Descriptive statistics, represented by bar charts of mean values, is used to explore the data. Wilcoxon Signed Ranks Test is used to find statistically significant differences between the variables in different scenarios. Concerns about sample size and its effect on statistical significance (see Section 7.2.7) still apply so statistical significance up until p-value of 0.1 is reported. The descriptive statistics for TFC departure variables is used to assess whether in the simulated environment there are still events of drivers not checking a signal aspect before departing from a station.

For departures, only scenarios 1 and 2 are compared. Scenario 3 is a complete copy of scenario 1 from the departure point of view.

### *8.3.6 Simulator's validity*

Albeit the widespread use of simulators in modern driving studies, much research (especially in rail industry) does not test the used simulators for real-life validity. Shechtman (2010), in his literature review, confirms that validation is necessary for simulator studies but there is no consensus on the means of validation between researchers. This study combines subjective and objective metrics to assess how the results of the experiment can be related to real-world and to explore the case for such simulator models.

The participants were asked to fill in a short questionnaire to assess the simulation. Appendix L includes a copy of this questionnaire. They were asked to rate such aspects of the simulated environment as visuals, train handling, immersion and comfort on 5-point Likert scale (very good to very poor). Moreover, the participating drivers were asked to assess the ease of noticing changes in physical environment and general validity of the simulation.

As data collection methodologies are similar for the in-field and simulator experiments, it is possible to compare drivers' performance based on the gaze data. Such approach has been used before to validate simulators. For example, Carter and Laya (1998) used eye-tracking to validate a fixed base car driving simulator and found no difference in spatial distribution of fixations on the visual field. Knapper *et al.* (2015) claim that the validity is affected by a task studied and population similarity. This was partially confirmed by significant effect of personal factors on many variables (Section 7.3.12). Hence only the data for the 2 drivers who participated in both experiments is compared. Moreover, data from scenario 3 is compared against the in-field data because the passenger loadings (deemed as major PSF) from that scenario better match the passenger volumes experienced during the in-field tests.

## 8.4 Results

### *8.4.1 Validity of the simulation environment*

The participants are unanimous in their assessment of the simulator returning five identical questionnaires with only comments and a suggestions box having different information. Table 29 below provides an overview of drivers' scores.

| Question | Answer |
|---|---|
| **Please rate visual aspect of the simulation** | Very good |
| **Please rate train handling aspect of the simulation** | Good |
| **Please rate immersion aspect of the simulation** | Very good |
| **Please rate comfort aspect of the simulation** | Very good |
| **Was it easy to notice changes introduced in the physical environment?** | Yes |
| **Would you consider this simulation as a good representation of driver's duties in T&W Metro?** | Yes |

**Table 29. Summary of participants' opinions on validity of the simulator**

According to the participants, the simulation is a good representation of real-life metrocar driving. However, the drivers struggled slightly with controls of the train hence the lower score for that question. The drivers state that they did notice changes between the scenarios.

Table 30 below shows results of statistical tests for difference between the real-life study and the simulation experiment (scenario 3) in terms of station arrivals.

| | **TFC** | | | | **AFD** | | | |
|---|---|---|---|---|---|---|---|---|
| | Felling | GStad | Heworth | Pelaw | Felling | GStad | Heworth | Pelaw |
| **Mirror** | 17.0 | 9.0 | 21.5 | 25.0 | 14.0 | .0** | 15.0 | 21.0 |
| **Platform** | 24.5 | 1.0** | 5.0** | 16.5 | 16.5 | 13.0 | 15.0 | 19.0 |
| **Signal** | 15.0 | 10.5 | 24.5 | 20.0 | 15.0 | 8.0** | 23.0 | 20.0 |
| **StopPos** | 22.5 | 19.0 | 24.5 | 23.0 | 4.0** | 11.0 | 11.0 | 9.0* |

**Table 30. Results of Wilcoxon Signed Rank Test looking into differences between real-life and simulated environment (arrivals) $*p<0.1$, $**p<0.05$**

88% (28 out of 32) of the variables do not show statistically significant difference between real life and simulation. Two differences are found in regards to platforms. Signals and mirrors have one significant difference each. When statistical significance is assumed at $p<0.1$ 84.4% of the variables do not show statistical significance. When tested against scenario 1 (lower passenger levels), the match between variables decreases to 69% with all eight platforms–related variables demonstrating statistically significant difference.

For the departure variables, the match is 89% (14 out of 16). The only two statistically significant differences are found for mirrors at Felling and Heworth (Table 31). Same match percentage is found for scenario 1. Even though there were concerns of how well the mirrors with clocks represent real-life DOO mirrors, the validity analysis showed that those provide good alternative.

| | **TFC** | | | | **AFD** | | | |
|---|---|---|---|---|---|---|---|---|
| | Felling | GStad | Heworth | Pelaw | Felling | GStad | Heworth | Pelaw |
| **Mirror** | 7.0** | 15.5 | 13.5 | 6.5** | 19.0 | 15.0 | 23.0 | 23.0 |
| **Signal** | 19.5 | 24.0 | 17.0 | 17.0 | 22.0 | 23.0 | 14.0 | 14.5 |

**Table 31.Results of Wilcoxon Signed Rank Test looking into differences between real-life and simulated environment (departures) *p<0.1; **p<0.05**

### *8.4.2 Influence of metro driving experience*

Nine test runs in three different scenarios were carried out by recently qualified drivers (under two years of experience) whereas five test runs were done by experienced drivers (driver coaches). The test runs are compared for all variables for three different metrics for both arrivals and departures. Table 32 below summarises results of Mann-Whitney U-test.

| | **Arrivals** | | | **Departures** | | |
|---|---|---|---|---|---|---|
| | TFD | TFC | AFD | TFD | TFC | AFD |
| **Scenario 1** | 0/16 | 1/16 | 0/16 | 1/8 | 1/8 | 1/8 |
| **Scenario 2** | 3/16 | 2/16 | 2/16 | 0/8 | 0/8 | 0/8 |
| **Scenario 3** | 2/16 | 1/16 | 4/16 | 0/8 | 1/8 | 0/8 |

**Table 32. Number of statistically significant differences in gaze behaviour (experienced vs novice drivers)**

### *8.4.3 Influence of design amendments during arrivals*

In order to see patterns in fixation durations, cumulative TFD values are calculated for each station for each scenario. Table 33 below summarises mean values of total fixation duration at all four design elements at each of the stations under investigation.

| | **Total fixation duration on 4 AOIs** | | | |
|---|---|---|---|---|
| | Felling | Gateshead Stadium | Heworth | Pelaw |
| **Baseline** | 8.2 | 7.76 | 6.45 | 7.88 |
| **Amended** | 7.2 | 7.12 | 7.12 | 7.47 |
| **More PAX** | 7.91 | 8.02 | 6.83 | 7.68 |

**Table 33. Mean cumulative TFD on all AOIs on arrivals**

According to table 33, the stations where spacing between the elements increased are the stations demonstrating a drop in TFD. Even though the signal at Felling is moved closer to other elements, its position on the other side of a track means that the element moves further apart in a visual field and the FOW angle increases as the participants get closer to a stopping point. On the contrary, Heworth, where the spacing is decreased and all the elements are on the same side, demonstrate increase in cumulative TFD.

When scenarios 1 and 3 are compared, one can note that more confined stations demonstrate higher cumulative TFD in scenario 3 whereas open stations return bigger values in scenario 1. As per Chapter 7, the confined stations are Heworth and Gateshead Stadium. Open stations are Pelaw and Felling. The only relationship which is close to statistical significance (*p*-value = 0.084) is between Gateshead Stadium scenarios 2 and 3.

Changes introduced at Pelaw mean that the AOIs are more spaced out, and the angle between the signal and the mirror is narrower. These changes result in visual attention shift from the elements responsible for stopping a train to the platform and the running signal in scenario 2 (Figure 40). Higher TFC/lower AFD cases can be observed for the mirror and the stopping position indicator (Figure 41 and Figure 42).

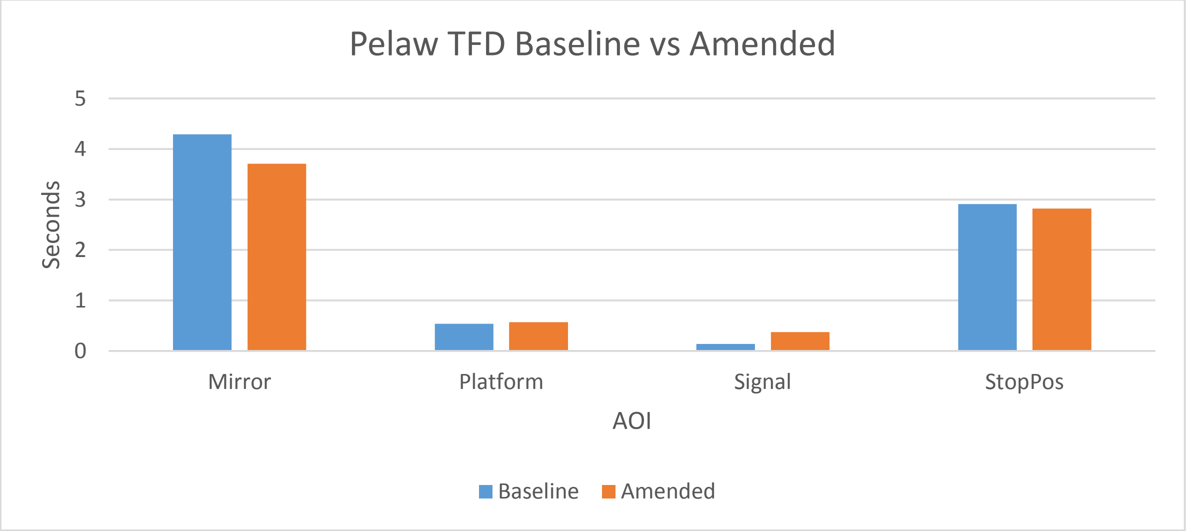


Figure 40. TFD statistics for Pelaw arrivals. Scenario 1 and 2

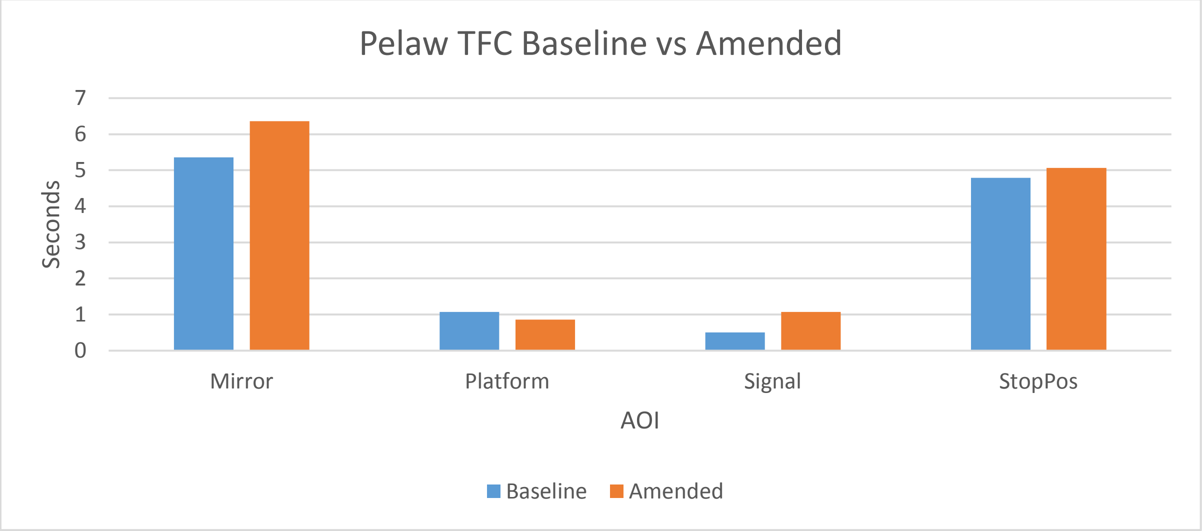


Figure 41. TFC statistics for Pelaw arrivals. Scenario 1 and 2

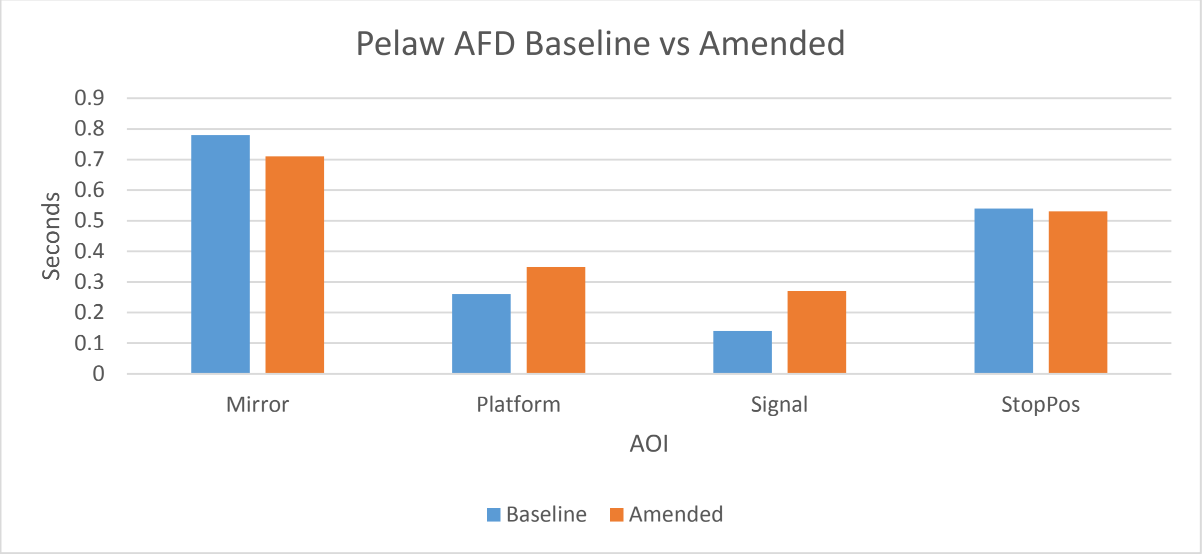


Figure 42. AFD statistics for Pelaw arrivals. Scenario 1 and 2

Changing the signal position at Heworth leads to reduced spacing between the AOIs and closer position of those in drivers' field of view. As all the elements are located on the same side of the track, an already narrow FOW angle becomes even smaller. As shown by Figure 43, moving a running signal closer to other elements at a side platform station increases TFD for a mirror and a signal. Marginal decrease is observed for gaze duration on a platform. A stopping position indicator shows more substantial decrease in TFD.

Changing the station layout does not cause any elements to exhibit higher TFC and lower AFD in scenario 2 (Figure 44 and Figure 45). However, one can observe that in comparison to the amended system, the platform and the stopping position indicator in the baseline layout show such trends.

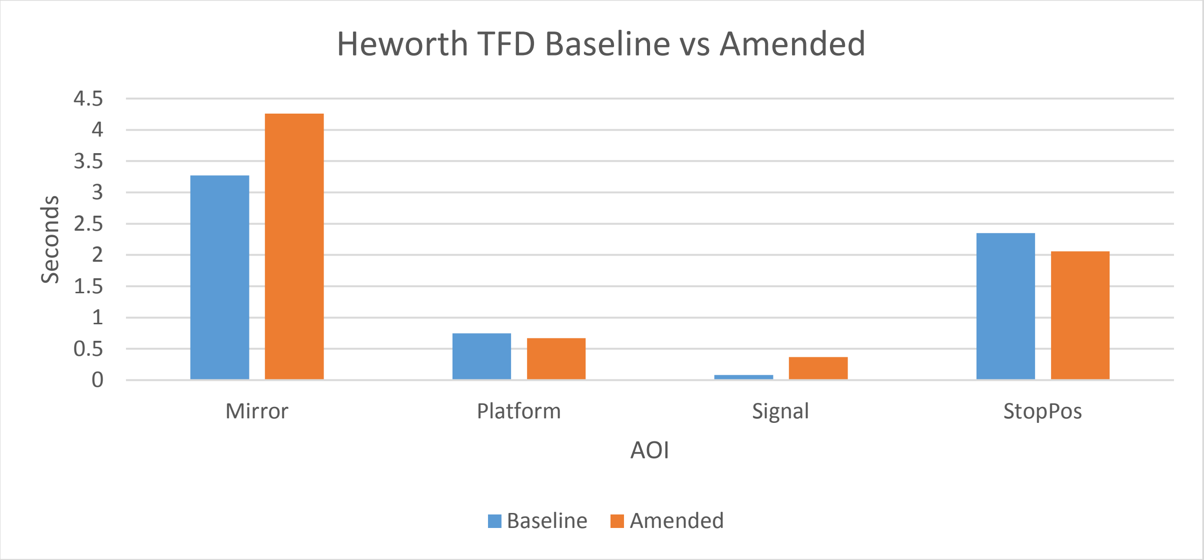


Figure 43. TFD statistics for Heworth arrivals. Scenario 1 and 2

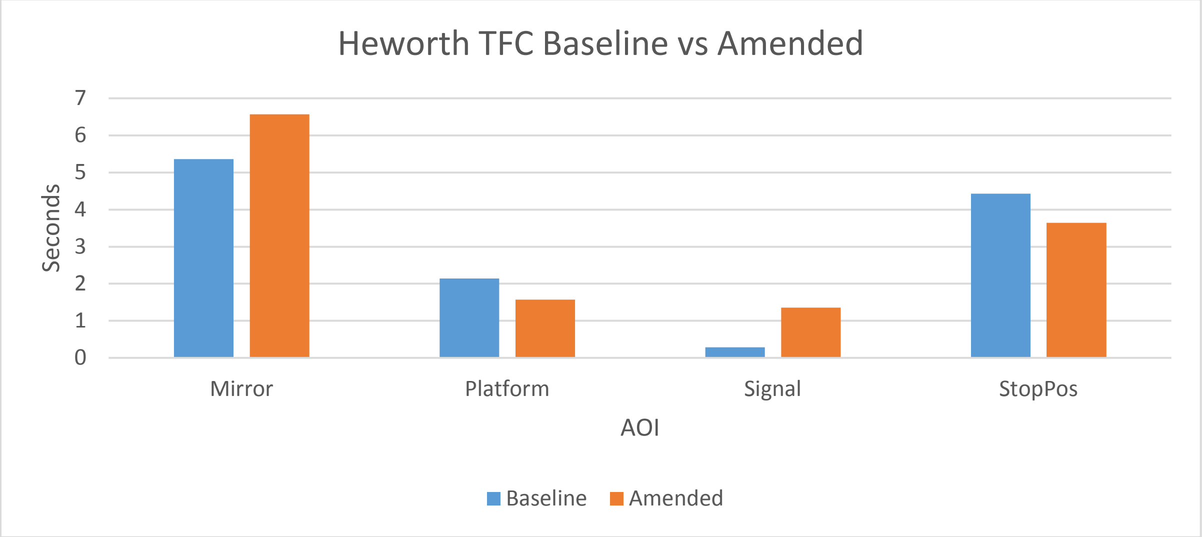


Figure 44. TFC statistics for Heworth arrivals. Scenario 1 and 2

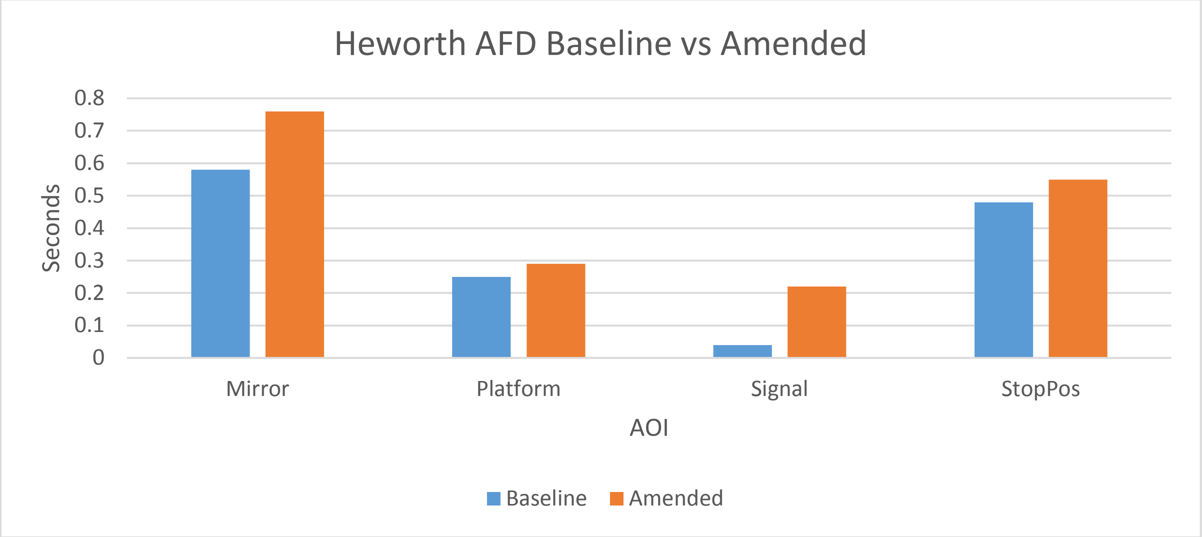


Figure 45. AFD statistics for Heworth arrivals. Scenario 1 and 2

The amended layout at Felling increases the angle between the signal and the mirror by bringing the signal closer to the platform edge. As with other stations, the introduced change causes increased TFD on Felling signal (Figure 46). On contrary, total fixation duration decreases for all other AOIs. All elements but the mirror demonstrate synchronised change in TFC and AFD metrics in the amended system (Figure 47 and Figure 48). The participants utilise more fixations, which are shorter, to interact with the mirror in scenario 2.

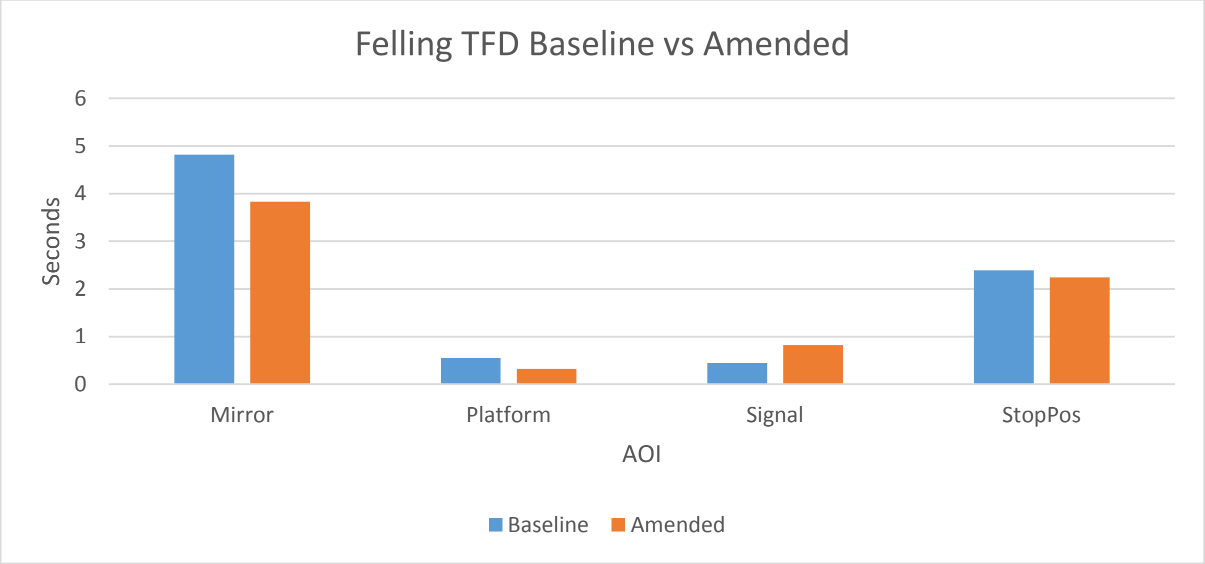


Figure 46. TFD statistics for Felling arrivals. Scenario 1 and 2

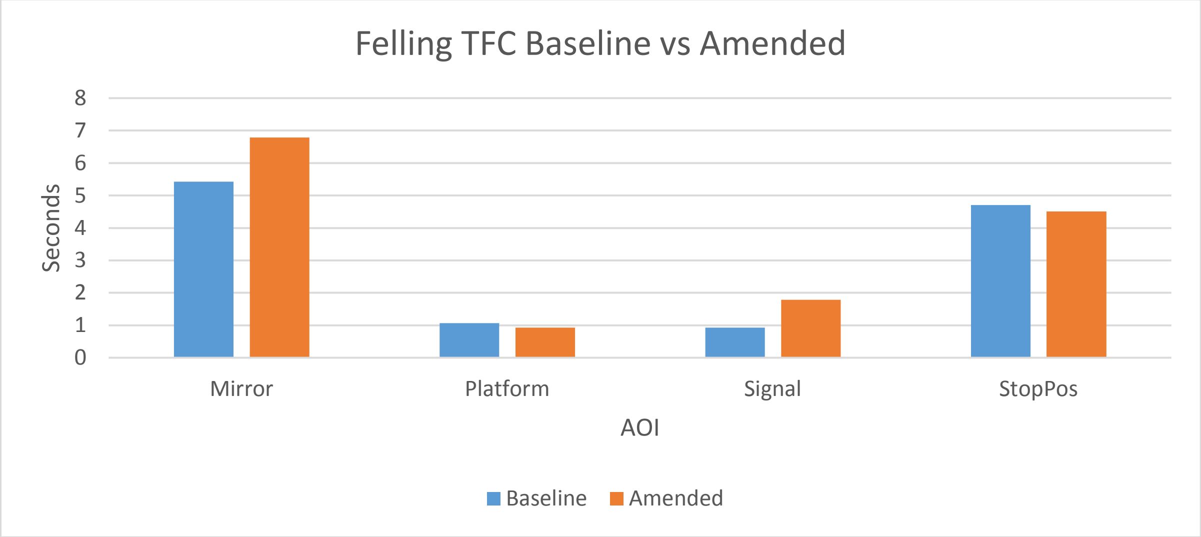


Figure 47. TFC statistics for Felling arrivals. Scenario 1 and 2

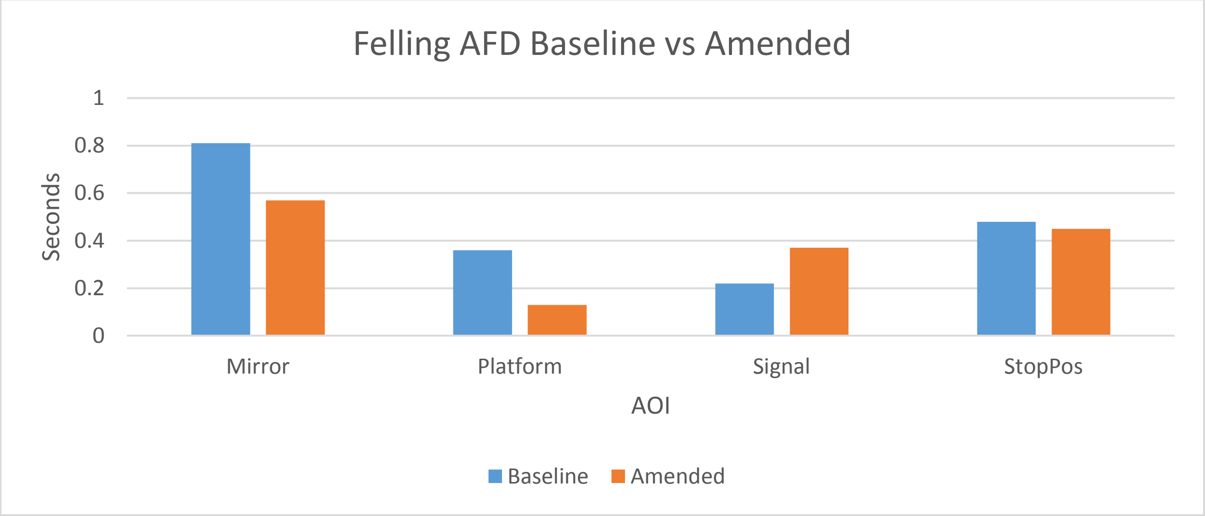


Figure 48. AFD statistics for Felling arrivals. Scenario 1 and 2

The running signal at Gateshead Stadium is moved further from the stopping position in scenario 2 thus reducing the angle between it, a driver, and the mirror. Moreover, the signal is located after an underbridge in scenario 2. The drivers look at it through a bridge arch. Figure 49 shows increase in TFD on the signal there, similarly to other stations. The platform maintains drivers’ attention on the same level whereas the mirror and the stopping position marker lose approximately 15%. Other stations where TFD on mirrors decreased showed more and shorter fixations on this element. However, the mirror at Gateshead Stadium demonstrates fewer and longer fixations in scenario 2 compared to the baseline scenario (Figure 50 and Figure 51). Similarly to all other locations, Gateshead Stadium signal shows growth in all three metrics. The results show higher TFC and lower AFD for the platform in scenario 2.

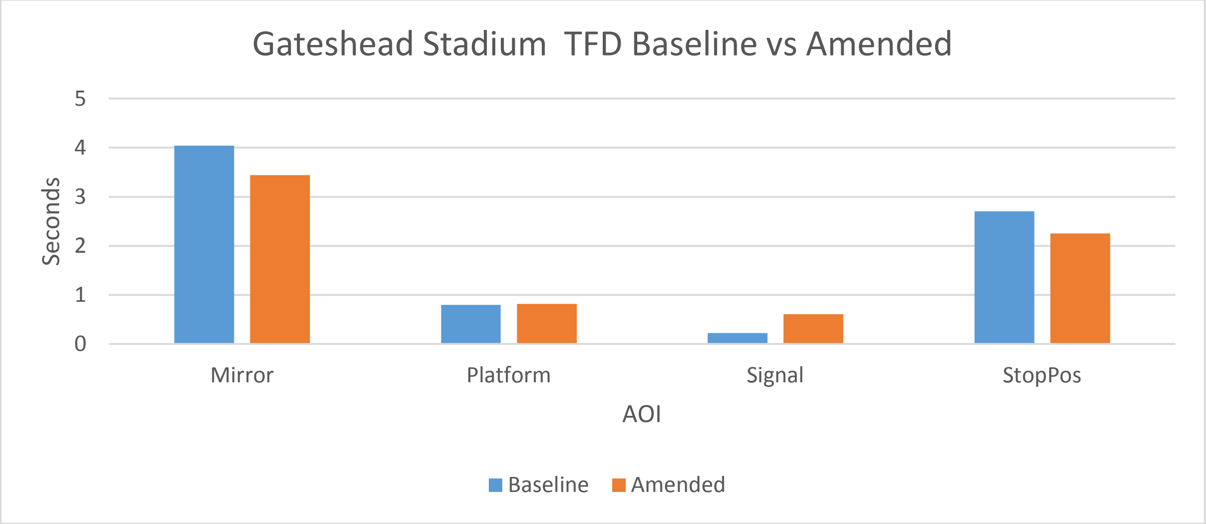


**Figure 49. TFD statistics for Gateshead Stadium arrivals. Scenario 1 and 2**

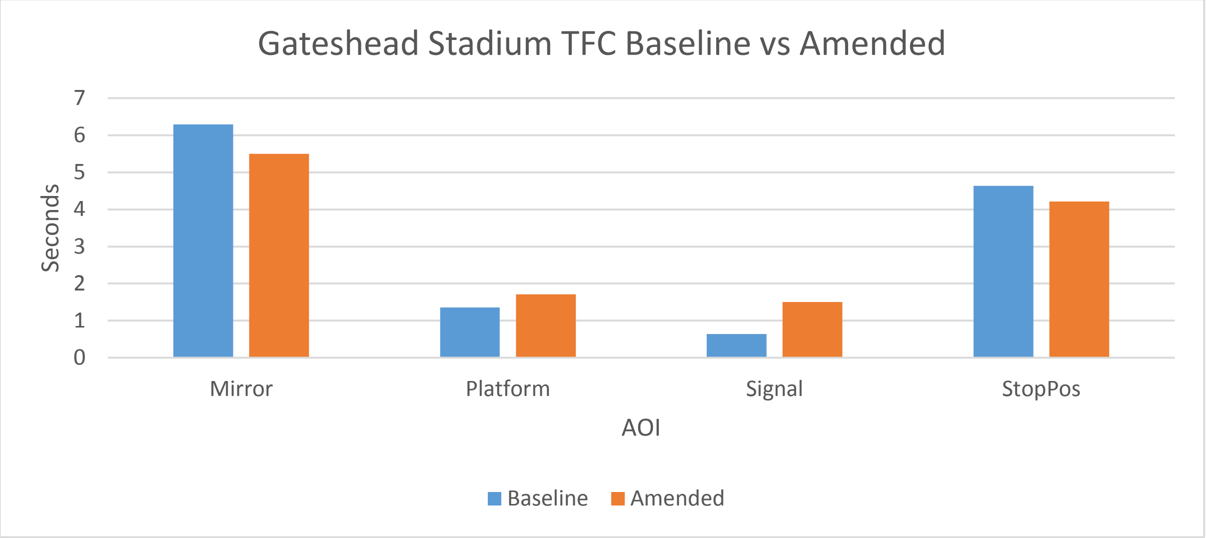


**Figure 50. TFC statistics for Gateshead Stadium arrivals. Scenario 1 and 2**

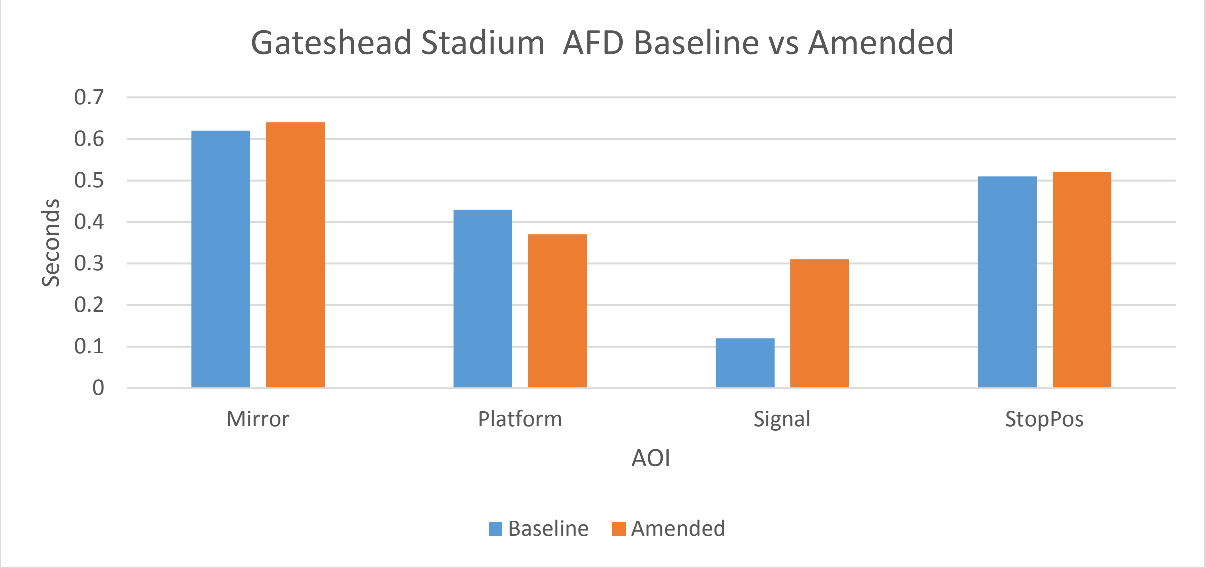


**Figure 51. AFD statistics for Gateshead Stadium arrivals. Scenario 1 and 2**

### *8.4.4 Influence of increased passenger loadings during arrivals*

Figure 52 below shows that increased passenger levels at Pelaw increase signal and platform TFD. The same metric for the mirror changes insignificantly but almost 30% decrease can be observed for the stopping position indicator. Figure 53 and Figure 54 demonstrate that the TFC and AFD for the platform and the signal at Pelaw change in line with TFD metrics. Higher TFC and lower AFD trend observed for the mirror in scenario 3. For the stopping position indicator, the number of fixations decreased but average fixation duration stayed almost the same. Magnitude of change between scenario 1 and 3 for TFC variables was considerably higher than for AFD variables at Pelaw.

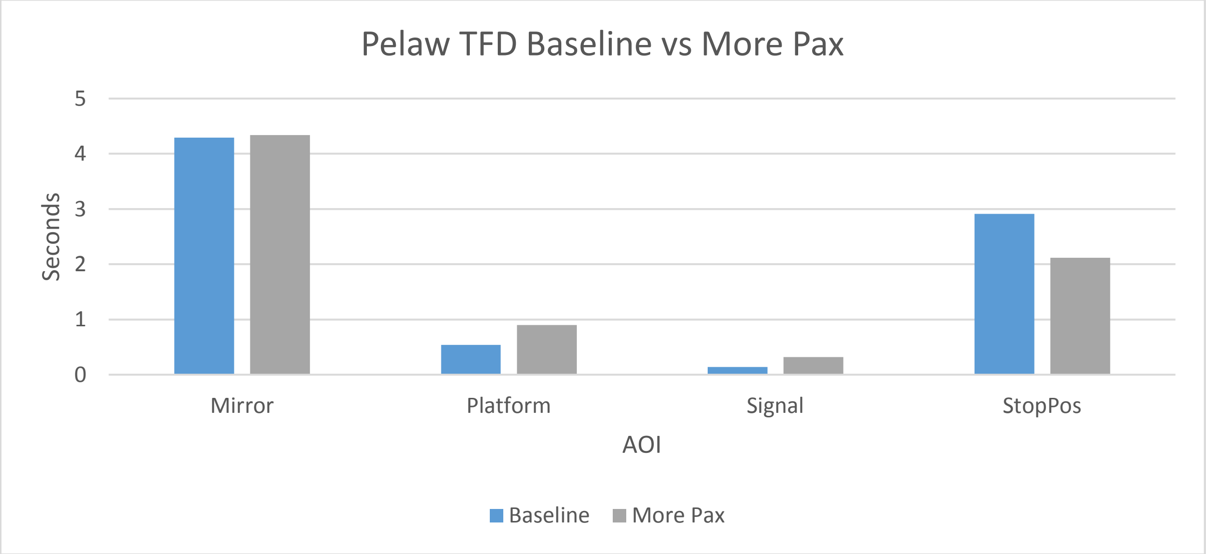


Figure 52. TFD statistics for Pelaw arrivals. Scenario 1 and 3

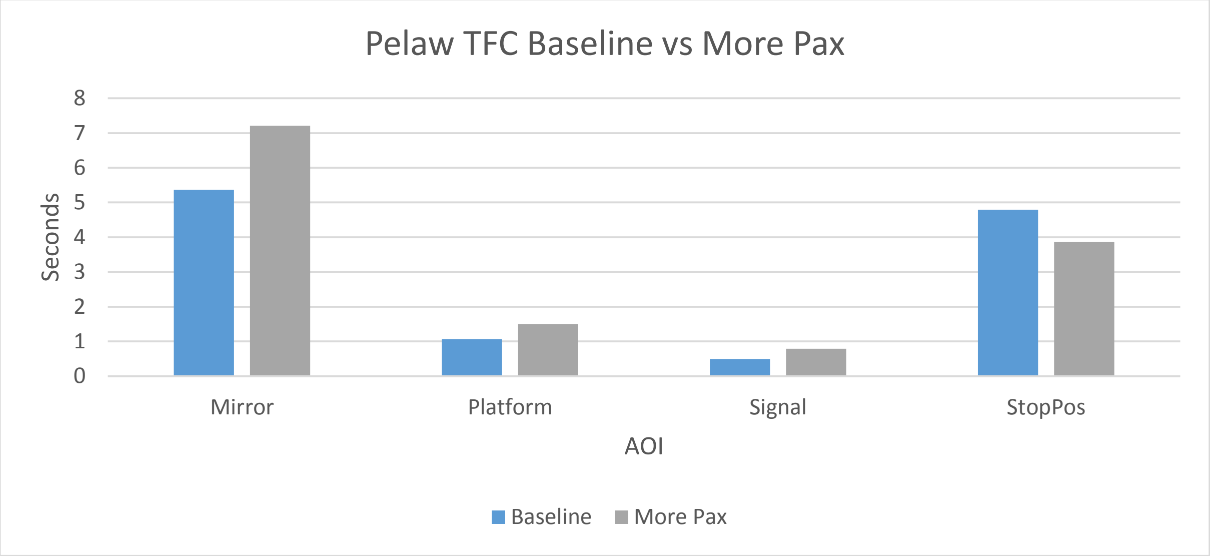


Figure 53. TFC statistics for Pelaw arrivals. Scenario 1 and 3

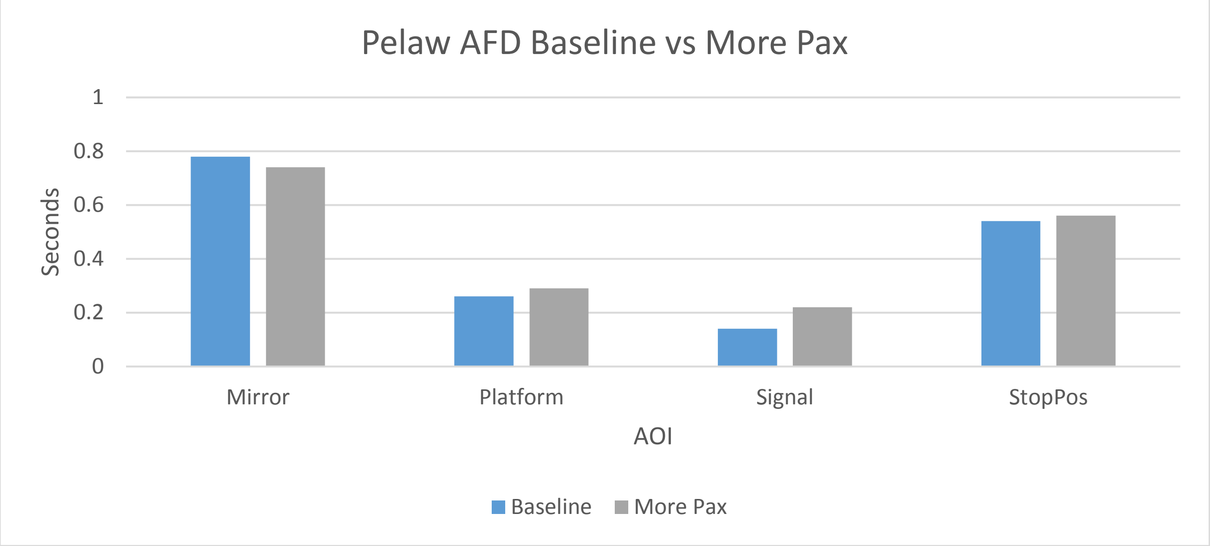


Figure 54. AFD statistics for Pelaw arrivals. Scenario 1 and 3

All the AOIs but the mirror demonstrate significant increase in TFD at Heworth in scenario 3 (Figure 55). According to Figure 56 and Figure 57, neither of the elements demonstrates trends implying problems with usability in high passenger loadings situation. Furthermore, the statistics shows that such trends can be observed for platforms in the baseline scenario however the magnitude of change for this element is rather small.

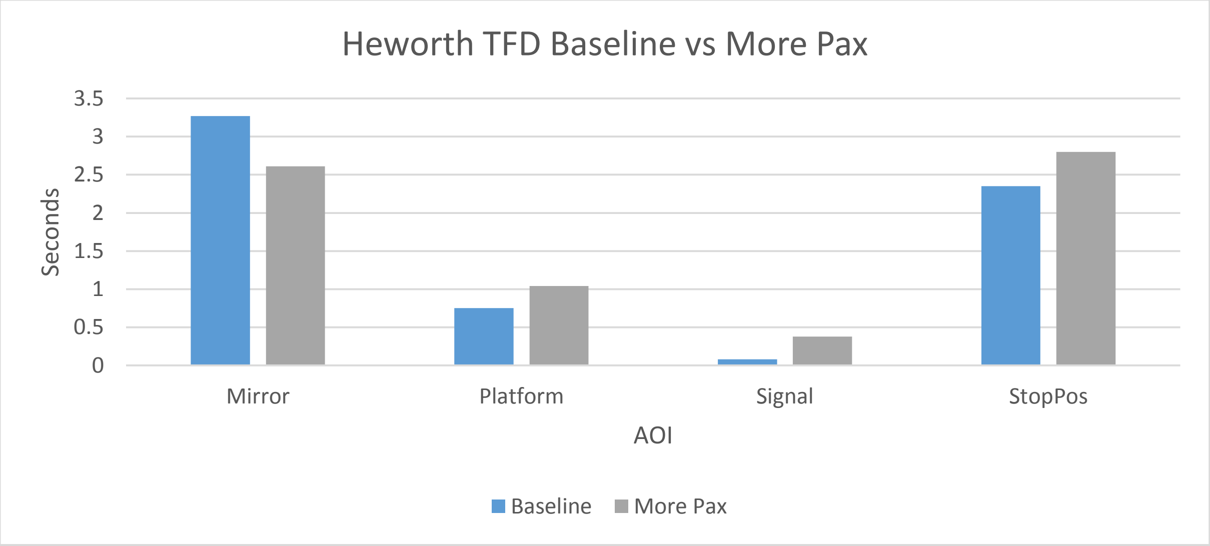


Figure 55. TFD statistics for Heworth arrivals. Scenario 1 and 3

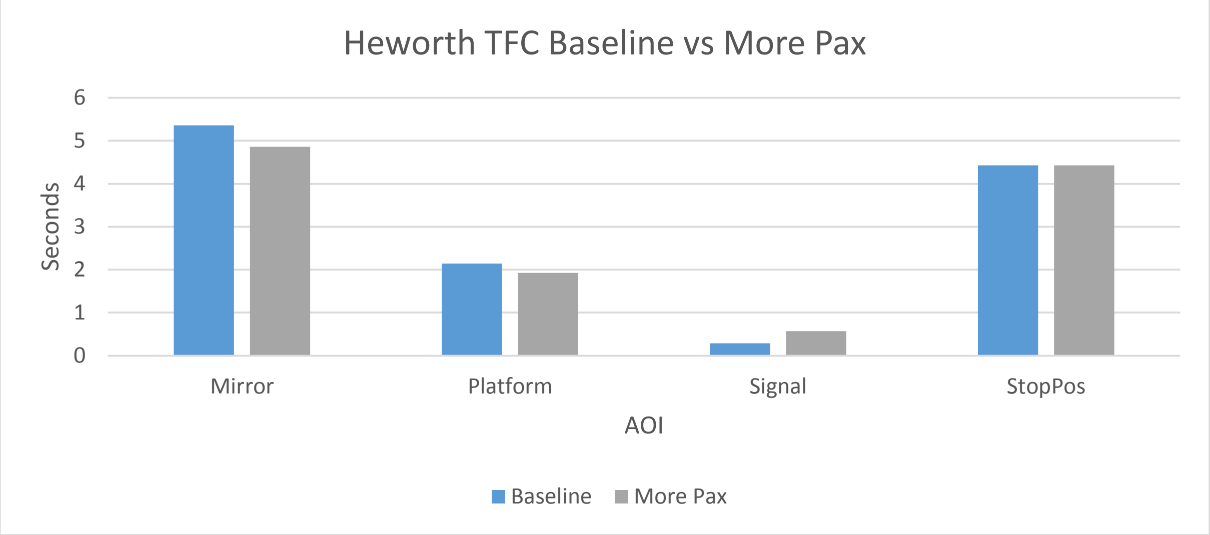


Figure 56. TFC statistics for Heworth arrivals. Scenario 1 and 3

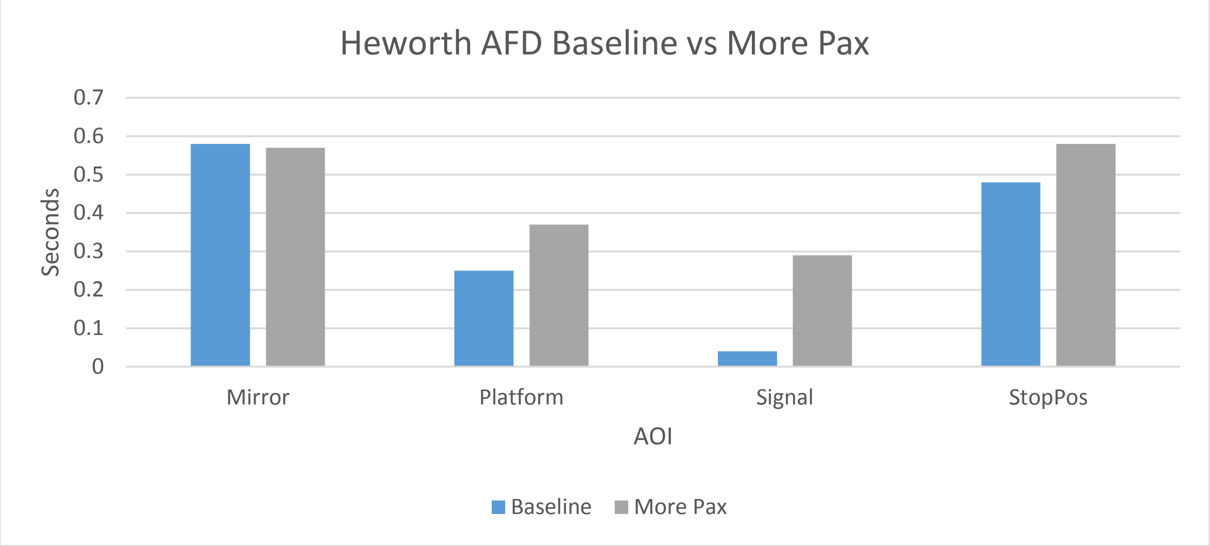


Figure 57. AFD statistics for Heworth arrivals. Scenario 1 and 3

Similarly to other stations, the drivers spend less time fixating on the mirrors (Figure 58) when there are more passengers at Felling. The platform demonstrates almost threefold increase in TFD. Felling is the only station where TFD on the running signal decreases in scenario 3. Finally, the stopping position marker also demonstrates a drop in total gaze duration on it.

As with other island stations, the drop in TFD on the mirror causes increase in fixation count and decrease in AFD (Figure 59 and Figure 60). The same trend can be observed for the stopping position marker. The other two AOIs demonstrate trends that are in line with TFD changes.

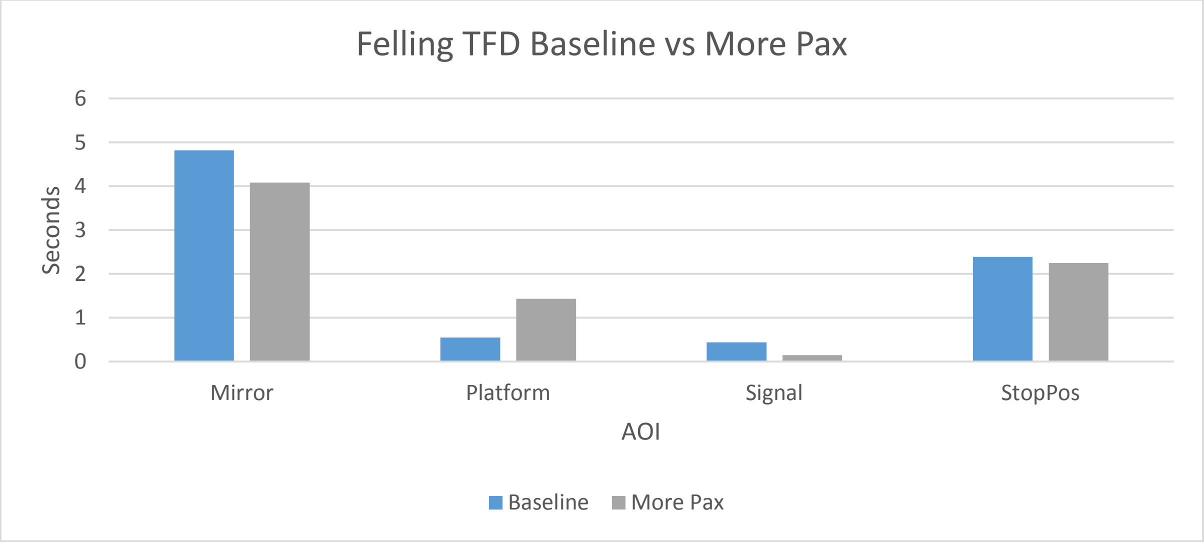


Figure 58. TFD statistics for Felling arrivals. Scenario 1 and 3

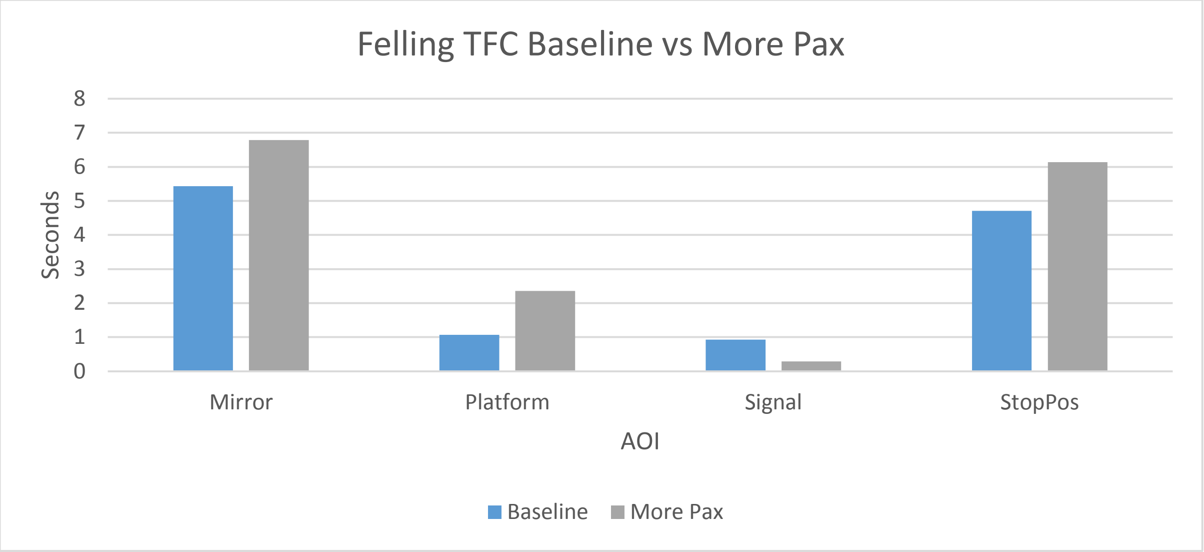


Figure 59. TFC statistics for Felling arrivals. Scenario 1 and 3

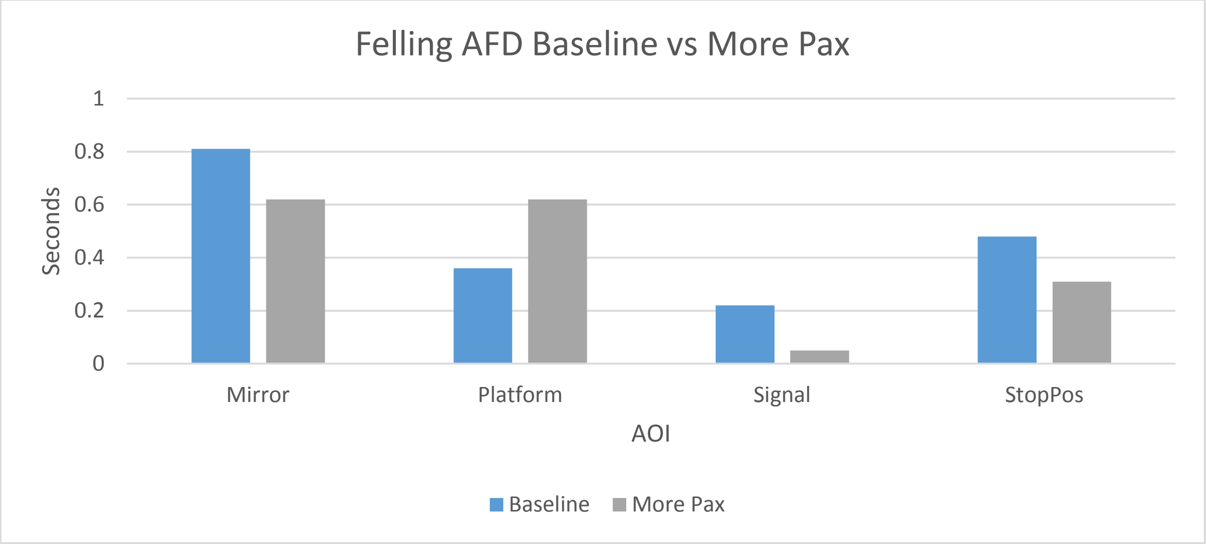


Figure 60. AFD statistics for Felling arrivals. Scenario 1 and 3

According to Figure 61, all of the AOIs but the mirror at Gateshead Stadium show increase in total fixation duration. A drop in TFD can be observed for the mirror. Figure 62 and Figure 63 highlight higher TFC/lower AFD trend for the mirror at Gateshead Stadium when passenger levels increase. The platform and the signal showed increase in both metrics whereas the stopping position marker only in TFC. Average fixation duration on the stopping position marker was the same in scenarios 1 and 3.

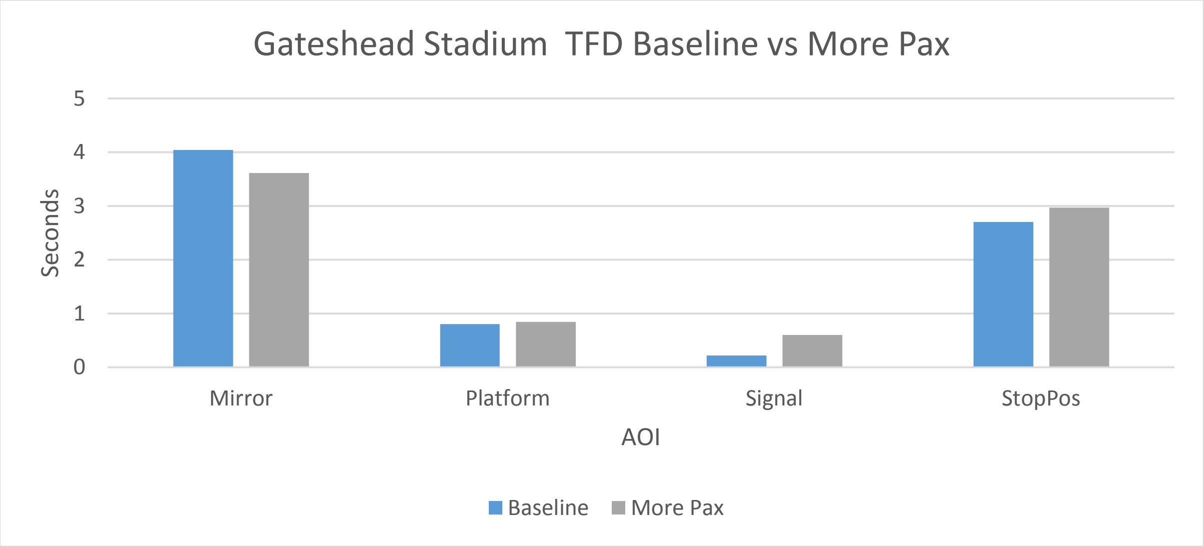


Figure 61. TFD statistics for Gateshead Stadium arrivals. Scenario 1 and 3

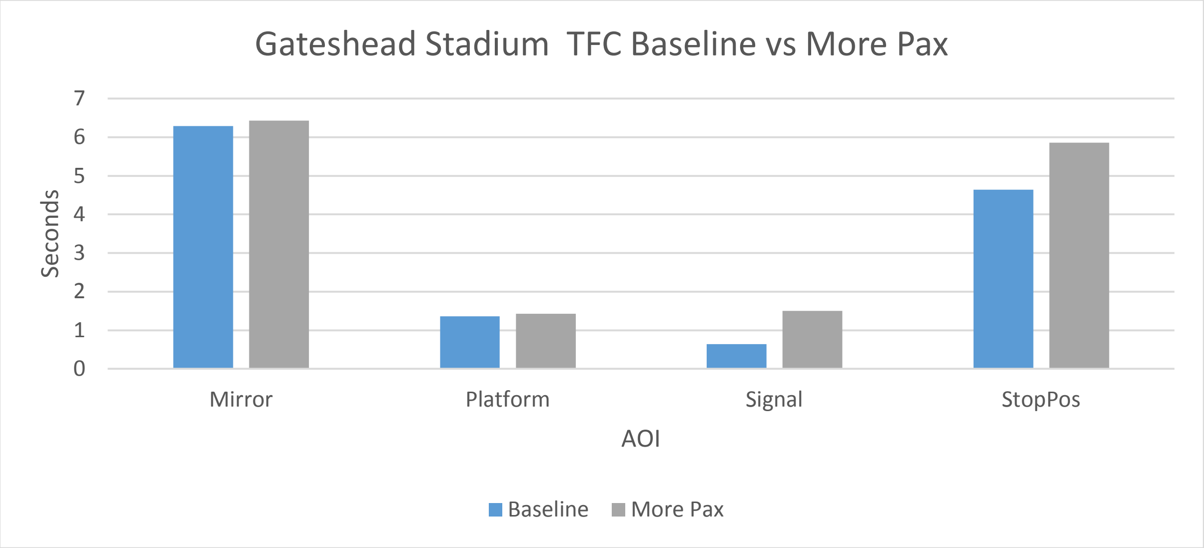


Figure 62. TFC statistics for Gateshead Stadium arrivals. Scenario 1 and 3

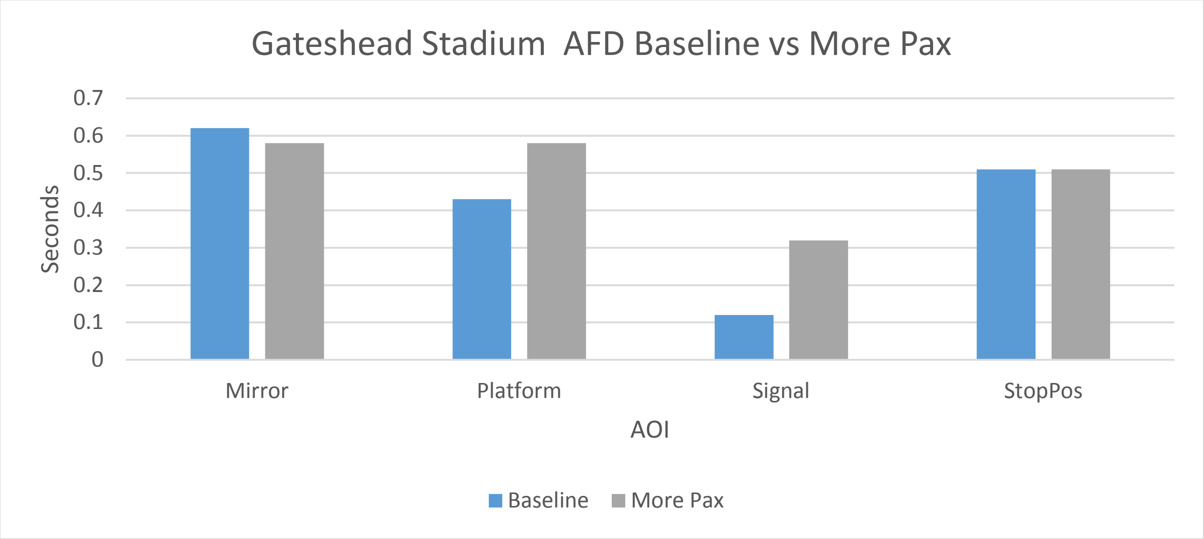


Figure 63. AFD statistics for Gateshead Stadium arrivals. Scenario 1 and 3

### *8.4.5 Influence of amended design during departures*

In order to see patterns in fixation durations, mean values are calculated for each station for scenarios 1 and 2. Table 34 below summarises mean values of cumulative TFD on all AOIs at each of the stations under investigation.

| | **Total fixation duration (mean)** | | | |
|---|---|---|---|---|
| | Felling | Gateshead Stadium | Heworth | Pelaw |
| **Scenario 1** | 7.54 | 7.52 | 6.78 | 7.05 |
| **Scenario 2** | 7.3 | 6.9 | 7.67 | 6.82 |

**Table 34. Mean cumulative TFD on all AOIs on departures**

All of the stations except of Heworth demonstrate graduate decrease in total dwell time on a signal and a mirror from scenario 1 to scenario 2. No statistically significant differences are found even when p-value range is expanded until 0.1.

As the figures below show (Figure 64, Figure 65, and Figure 66) the slight decrease in total fixation durations at Pelaw in scenario 2 is attributable to both AOIs. The mirror shows increase in TFC and decrease in AFD. The signal shows a minor increase in both metrics.

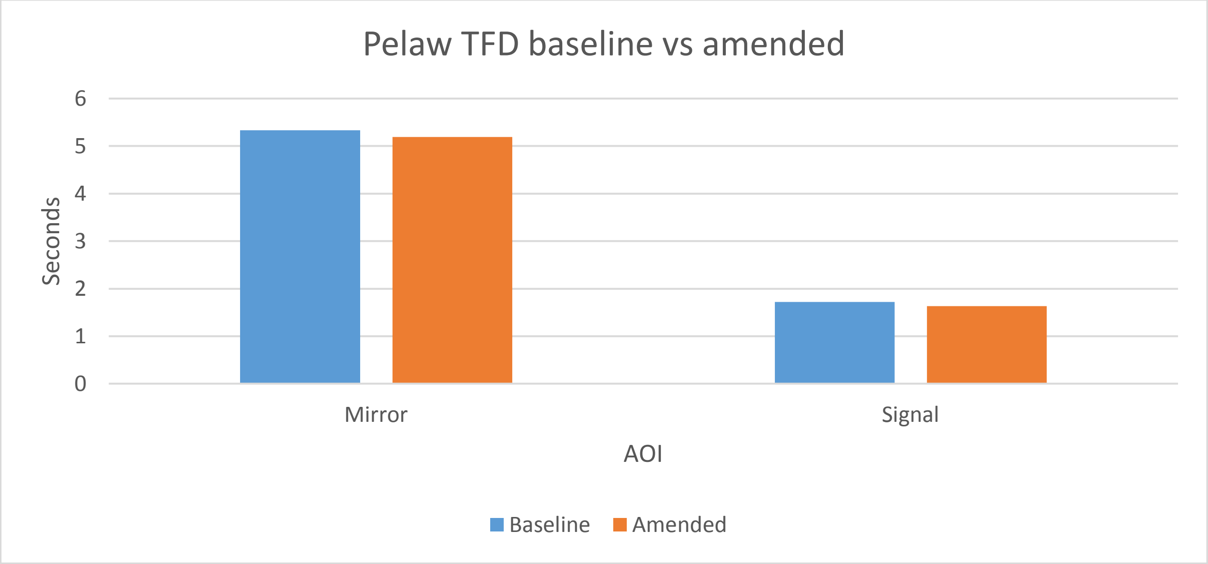


**Figure 64. TFD statistics for Pelaw departures. Scenario 1 and 2**

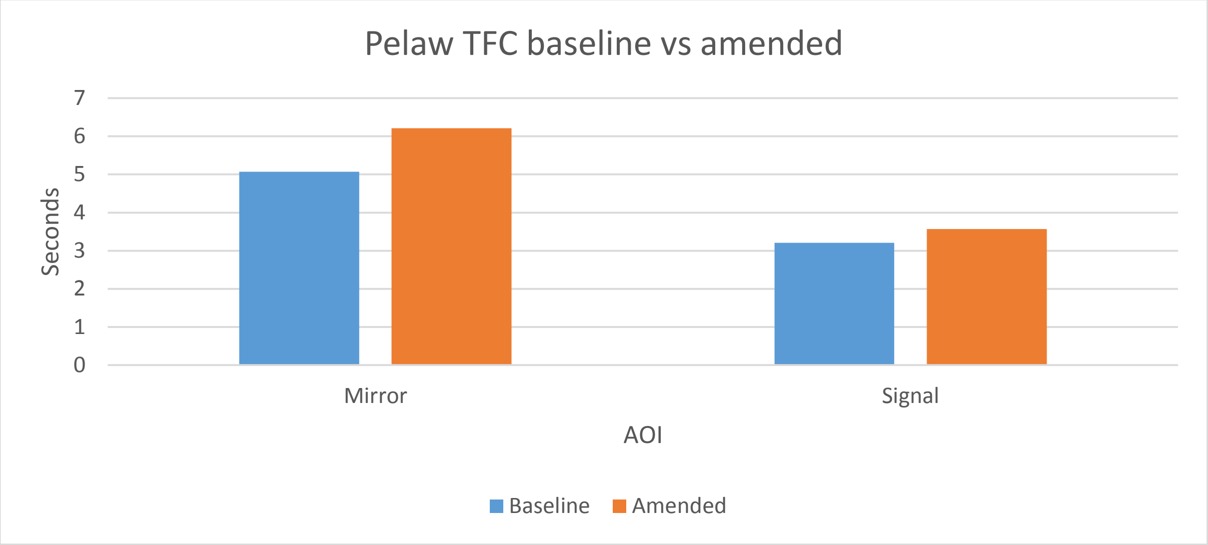


**Figure 65. TFC statistics for Pelaw departures. Scenario 1 and 2**

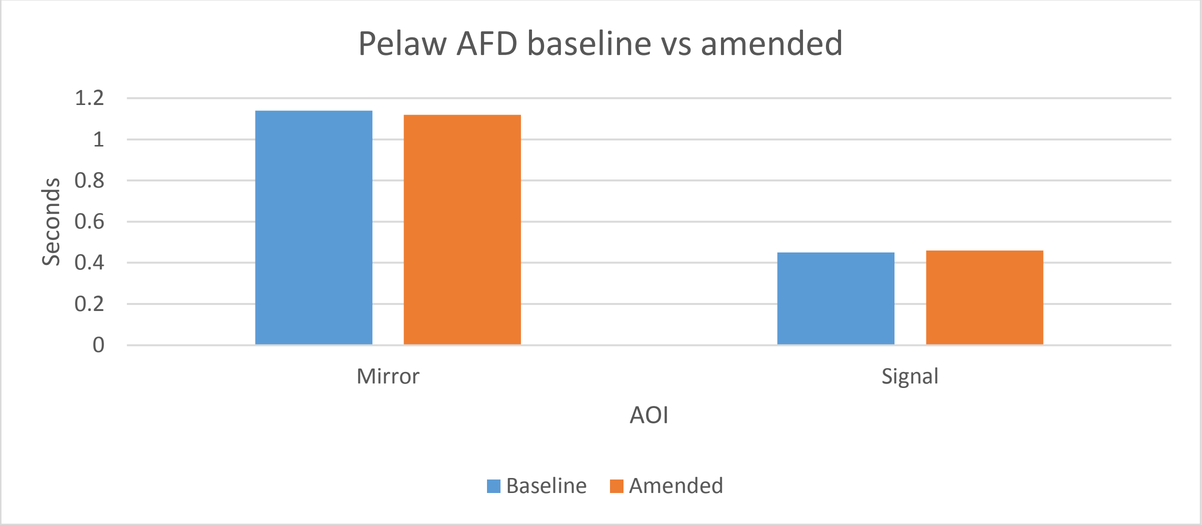


**Figure 66. AFD statistics for Pelaw departures. Scenario 1 and 2**

Change of the signal position at Heworth causes increased gaze durations on the signal and a minor drop for the mirror (Figure 67). This drop, however, is associated with a trend, which suggests drivers experiencing higher workload and stress when looking on the element (Figure 68 and Figure 69).

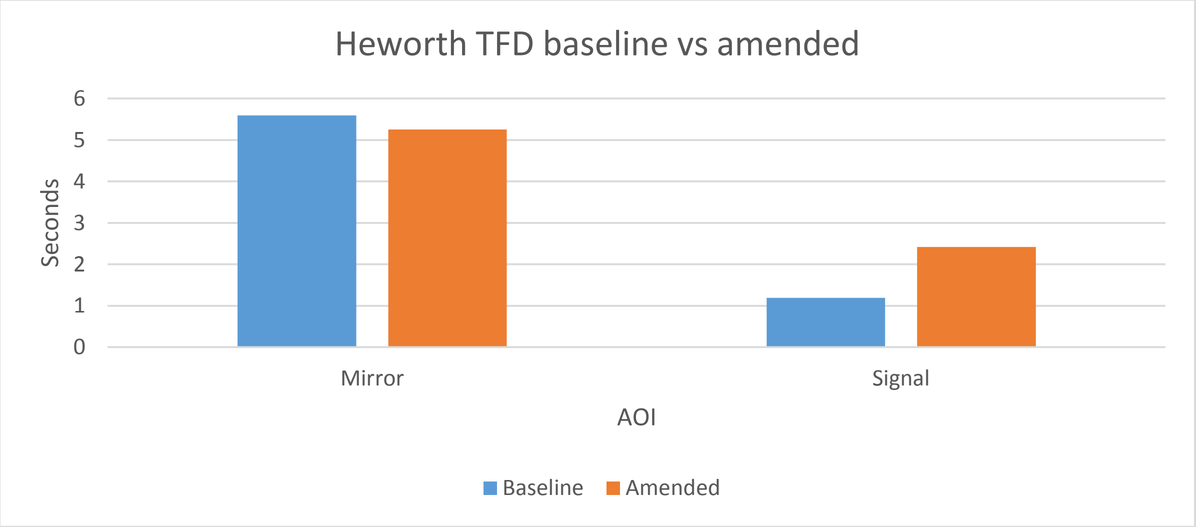


Figure 67. TFD statistics for Heworth departures. Scenario 1 and 2

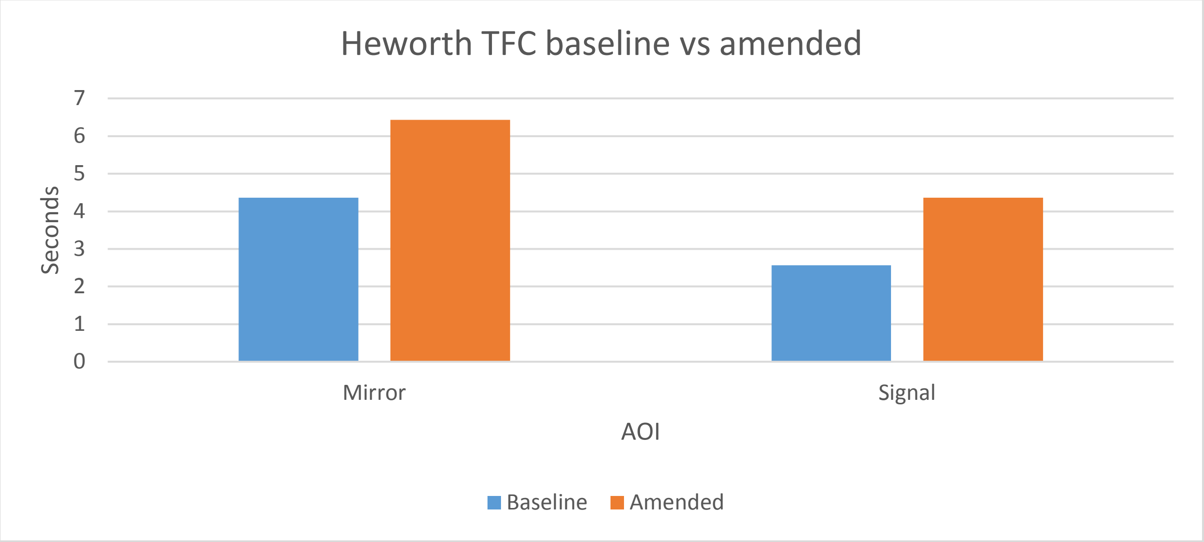


Figure 68.TFC statistics for Heworth departures. Scenario 1 and 2

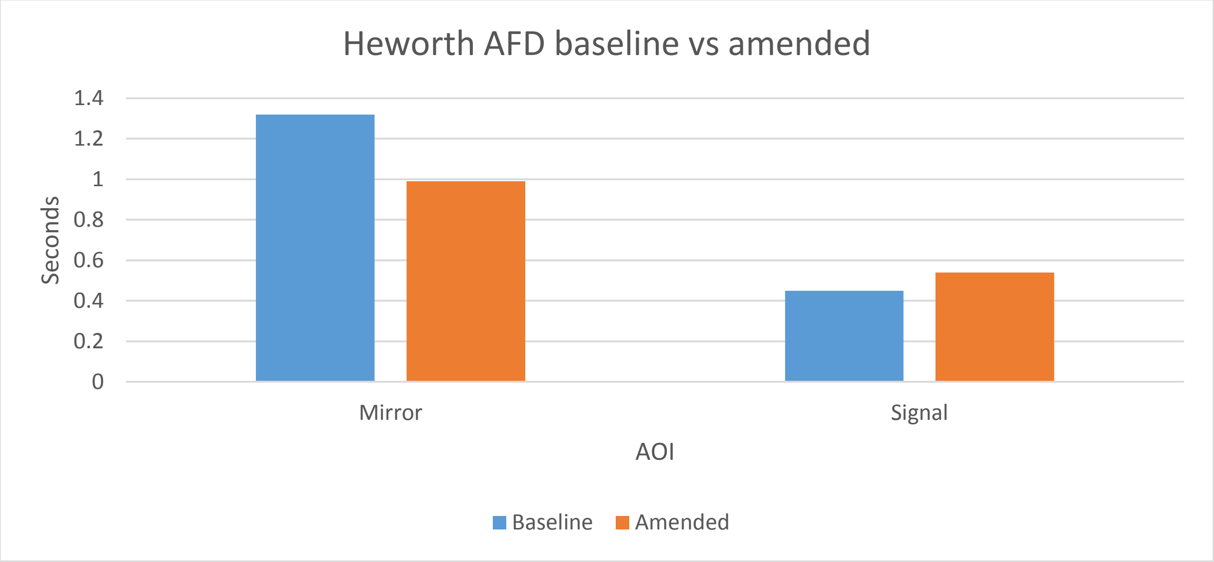


Figure 69. AFD statistics for Heworth departures. Scenario 1 and 2

Trends similar to Heworth are observed at Felling, where the running signal is moved closer to a stopping position (Figure 70, Figure 71, and Figure 72). However, TFD of Felling signal changes considerably less compared to the same metric for Heworth signal.

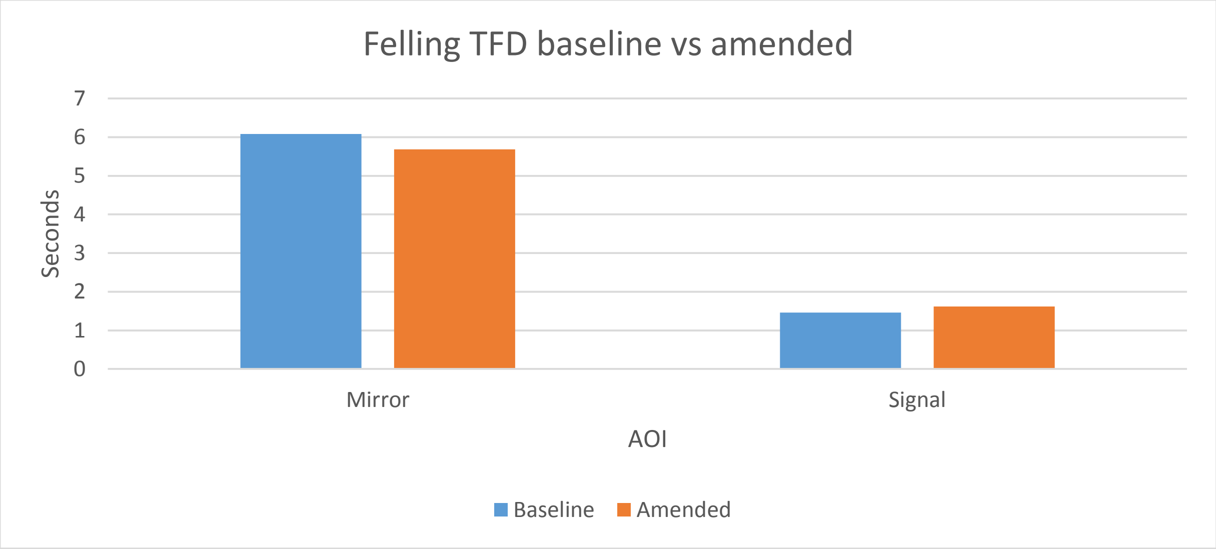


Figure 70. TFD statistics for Felling departures. Scenario 1 and 2

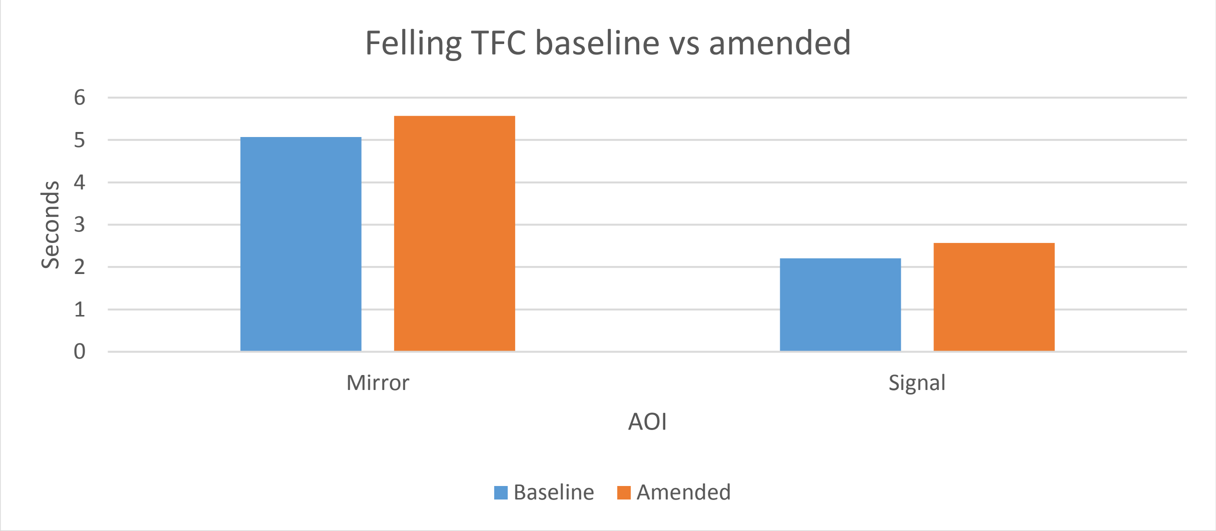


Figure 71. TFC statistics for Felling departures. Scenario 1 and 2

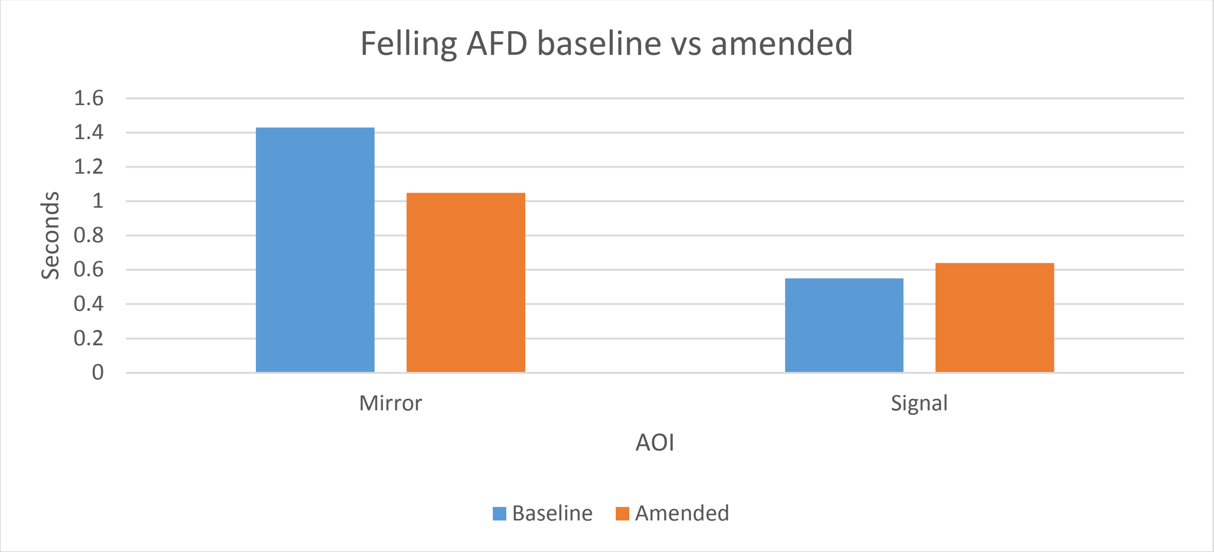


Figure 72. AFD statistics for Felling departures. Scenario 1 and 2

When Gateshead Stadium signal is moved further away, the observed change in gaze performance is similar to Heworth and Felling. The magnitudes of change are closer to Heworth than to Felling for TFD on the signal, but TFD on the mirror drops considerably more than anywhere else (Figure 73). Similarly, the AFD and TFC statistics show that drivers employ more and shorter fixations to interact with the mirror at Gateshead Stadium (Figure 74 and Figure 75).

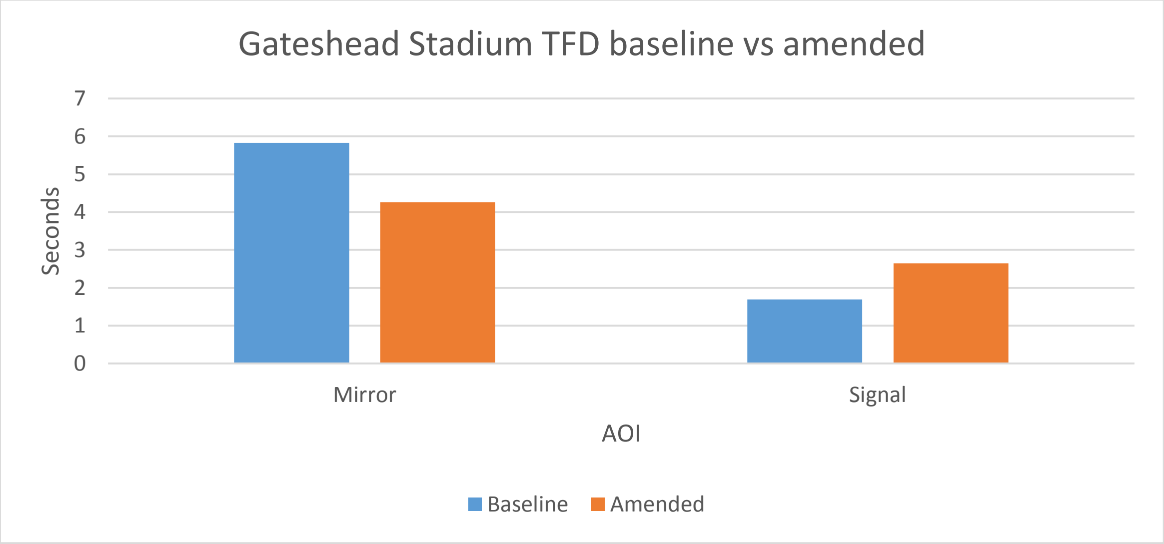


**Figure 73. TFD statistics for Gateshead Stadium departures. Scenario 1 and 2**

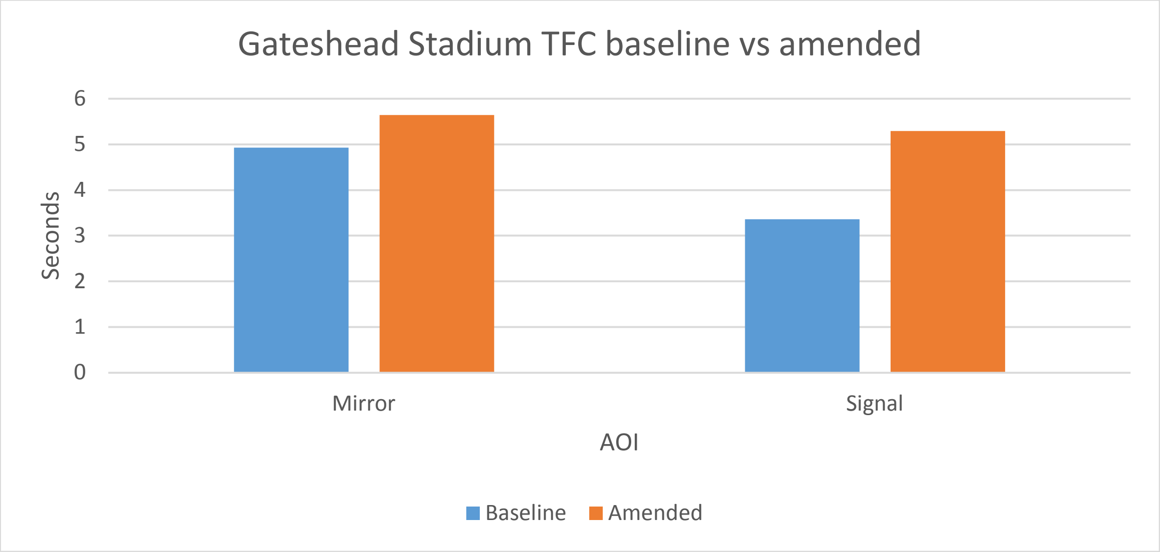


**Figure 74. TFC statistics for Gateshead Stadium departures. Scenario 1 and 2**

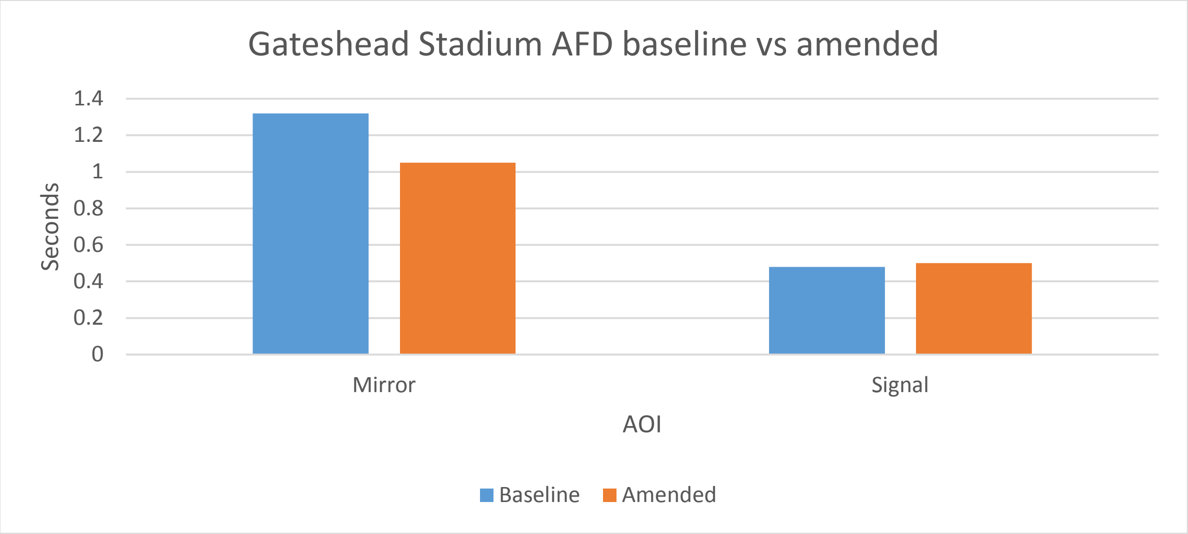


**Figure 75. AFD statistics for Gateshead Stadium departures. Scenario 1 and 2**

### 8.4.6 Failure to check a signal on departure

The real-life eye-tracking study has highlighted an issue of drivers' potentially violating an established rule book procedure for departures, namely not checking a signal (Section 7.4.5). Table 35 provides the results for the simulation study in terms of drivers' failing to check the signals on departure. It is important to note, that only one miss was created through the missed values analysis. It happened at Pelaw in scenario 3.

| | Scenario 1 | Scenario 2 | Scenario 3 | Real Life |
|---|---|---|---|---|
| **Number of tests** | 14 | 14 | 14 | 20 |
| **Pelaw** | 2 | 1 | 2 | 5 |
| **Heworth** | 0 | 0 | 1 | 2 |
| **Felling** | 2 | 0 | 1 | 3 |
| **Gateshead Stadium** | 1 | 0 | 0 | 1 |
| **Average per test** | 0.36 | 0.07 | 0.29 | 0.55 |

**Table 35. Summary of departures where drivers did not check a signal**

### 8.4.7 Results of the statistical tests

Wilcoxon Signed Ranks Test was used to perform statistical analysis to reveal differences between the variables in different scenarios. Table 36 below provides an overview of Z-values for arrival variables based on negative ranks. Only 17% of variables are statistically different between scenarios 1 and 2. Majority of the differences are for signal-related variables. Same trends are observed in comparison between scenarios 1 and 3. Another 9% of the relationships might return statistically significant differences in a bigger sample due to having *p*-value under 0.1. This proportion is similar to the in-field experiment.

| | | **TFD** | | **TFC** | | **AFD** | |
|---|---|---|---|---|---|---|---|
| | AOI | Scen. 2 | Scen. 3 | Scen. 2 | Scen. 3 | Scen. 2 | Scen. 3 |
| **Pelaw** | Mirror | -.722 | -.094 | -1.127 | -1.030 | -.245 | -.502 |
| | Platform | -.059 | -.866 | -.866 | -.483 | -.415 | -.357 |
| | Signal | -2.293** | -1.660* | -2.126** | -1.414 | -1.689* | -1.020 |
| | Stop.Pos. | -.489 | -1.287 | -.176 | -1.020 | -.549 | -.157 |
| **Heworth** | Mirror | -1.664* | -1.350 | -1.069 | -.433 | -1.256 | -.105 |
| | Platform | -.549 | -1.350 | -1.015 | -.356 | -.624 | -.974 |
| | Signal | -2.491** | -2.547** | -2.877** | -1.633* | -2.937** | -2.197** |
| | Stop.Pos. | -.664 | -1.538 | -.749 | -.282 | -.734 | -1.351 |
| **Felling** | Mirror | -1.099 | -1.244 | -2.189** | -2.430** | -1.852* | -1.853* |
| | Platform | -.785 | -1.726* | -.503 | -1.709* | -1.020 | -.784 |
| | Signal | -1.601 | -1.400 | -1.630* | -1.638* | -1.157 | -1.992** |
| | Stop.Pos. | -.594 | -.408 | -.155 | -1.386 | -1.119 | -2.544** |
| **G. Stad** | Mirror | -.157 | -.220 | -.777 | -.672 | -.560 | -.283 |
| | Platform | -.220 | -.157 | -.914 | .000 | -.377 | .000 |
| | Signal | -2.203** | -2.188** | -1.723* | -2.203** | -2.316** | -2.432** |
| | Stop.Pos. | -1.223 | .000 | -1.027 | -1.513 | -.105 | -.031 |

**Table 36. Results of Wilcoxon Signer Ranks Test comparing arrival AOIs in scenario 1 to scenarios 2 and 3. * p<0.1, ** p<0.05**

Table 37 below provides an overview of Z-values for departure variables based on negative ranks. Only 21% of variables are statistically different between scenarios 1 and 2. Majority of the differences are for signal-related variables. One more relationship could have returned statistically significant differences in a bigger sample due to having *p*-value under 0.1.

| | | **TFD** | **TFC** | **AFD** |
|---|---|---|---|---|
| | AOI | Scenario 2 | Scenario 2 | Scenario 2 |
| **Pelaw** | Mirror | -.094 | -1.316 | -.722 |
| | Signal | -.384 | -.877 | -.035 |
| **Heworth** | Mirror | -.722 | -.2683** | -1.601 |
| | Signal | -2.354** | -2.520** | -.596 |
| **Felling** | Mirror | -1.161 | -.994 | -1.538 |
| | Signal | -.594 | -.881 | -1.099 |
| **Gateshead Stadium** | Mirror | -1.977** | -.358 | -1.475 |
| | Signal | -1.852* | -2.123** | -.094 |

**Table 37. Results of Wilcoxon Signer Ranks Test comparing departure AOIs in scenario 1 to scenarios 2. * p<0.1, ** p<0.05**

## 8.5 Discussion

### *8.5.1 Categorisation of the developed simulator*

The final product developed to simulate virtual reality combines features of high level simulators and trade-offs of low level systems. Albeit having no immediate physical environment, the drivers are able to interact with the cab in the virtual world. Lack of moving base, as shown by literature review, should not affect drivers' perceptions of proposed scenarios. Similarly, using a flat screen can have a positive effect on the validity. In terms of motivational pressures, previous rail simulator studies showed that this parameter of a simulation study is rather realistic in virtual environments (Van Luipen and Meijer, 2009; Yates and Sharples, 2009). The railway systems normally have a small number of simultaneous events happening around a driver compared to road network thus addressing an issue of complexity of building a realistic environment raised by Shechtman (2010). Taking all of the above into account, it is possible to claim that the simulator used in this research is high level. This is supported by part-task trainers, for example the one built by Rowe (2013), being considered as medium fidelity installations.

The developed simulator demonstrates that new VR technologies and novel human-machine interfaces allow building a high level product in short timeframe and with a budget of under £1000. However, the validity of it should be assessed too. It is important to note that further advancements in VR technologies can make high level simulators obsolete due to better cost-benefit ratios and flexibility of the final product. Train Simulator 2016 already supports VR glasses, e.g. Oculus Rift, making the need for immediate physical environment irrelevant. This is one of the reasons why the current categorisation will have to be reconsidered.

### *8.5.2 Simulator validity*

Both subjective and objective measures confirmed high degree of match between the real-life and proposed VR train driving. Using eye-tracking measures on key areas of interest is a novel approach to validate a rail simulator. Even the previous eye-tracking validations in automotive industry defined AOIs in the immediate physical environment (car interior). Hence this validation approach provides even higher degree of confidence in representation of real-life metro driving. Fluctuations in the number of statistically significant differences with a change in passenger volumes not only confirm significant influence of this factor but also provide an important lesson in design of virtual environments. It is necessary to re-create the scenarios as much as possible to real-life conditions in order to gain meaningful results, which has been demonstrated before (Bella, 2005; Yan *et al.*, 2008; Santiago-Chaparro *et al.*, 2011). Moreover, this change also confirms that the simulator creates good motivational pressures.

The simulation methodology chosen shows how modern technologies allow building cheap high fidelity simulators. Additional advantage of this approach is a user-friendly interface that enables traincrew managers and coaches easily altering the virtual reality to simulate different situations. The biggest uncertainty associated with this method is what the user agreement for Train Simulator 2016 says about using the software for commercial purposes (traincrew training). However, making the custom-made VR railway systems publicly available (open access) could potentially resolve this issue.

The arrival statistics also shows that drivers spent more than 50% of the arrival time looking outside of the AOIs, most likely controls. This is a considerable increase compared to the in-field results. The same increase has not been observed in departure sequences meaning that controlling train motion might be the cause of it. As the drivers rated the train’s physics very high, it can only mean that the controller itself was at fault. It is not similar to the real-life master controller in any way. Hence tactile memory cannot be initiated and the actions require visual confirmation. In the future using a flight simulator controller should be considered if full VR set-ups are not available by then.

Despite some limitations, the results demonstrate that the simulator is a good tool to study relative, if not absolute, effects of changes in physical environment on drivers' performance. Moreover, the simulators allow pro-active assessment and modelling of infrastructure can bring significant cost and safety benefits to railways (Aitchinson and Davies, 2009). This thesis corroborates views of Simmons (2015) who advocates use of simulation and virtual world for "designing out" adverse PSFs.

### *8.5.3 Influence of metro driving experience*

With only 10% of arrival variables across all scenarios and 6% of the departure variables being influenced by experience level, it is possible to claim that this is not an important factor thus not confirming hypothesis 8e. For the arrivals, it is possible to see that all the drivers interact with the physical environment in the same way when driving in the original scenario. As it was suggested previously, the route knowledge of metro drivers is very high even for unexperienced drivers due to the size and type of a system. This corroborates the findings from RSSB (2005) report.

The difference becomes more pronounced in the amended scenarios where almost 17% of the variables show dependence on driving experience. It is known for less experienced participants to utilise different, potentially not as sophisticated, visual strategies (Reingold *et al.*, 2001; Savelsbergh *et al.*, 2002; Reingold and Charness, 2005). As the results show, the drivers might differ in terms of their preparedness for unexpected events depending on their driving experience. For example, in scenario 3 the experienced drivers have longer gazes on mirrors than the inexperienced colleagues who potentially get more distracted by increased passenger levels. The other trends observed demonstrate better prioritisation by the experienced group. These results support earlier results on SPaDs (Section 5.3.5), which demonstrated that experience level and presence of a disruption are important PSFs.

### *8.5.4 Fixation time on all AOIs*

Even though fixations are distributed differently across the scenarios, total fixation time on all AOIs is statistically not different for all stations. Some stations demonstrate approximately 13% differences between the scenarios but are deemed statistically similar by the selected tests. This controversy is further highlighted when individual variables are compared between the scenarios and only 17% found to be statistically different. This is however in line with the results from the in-field eye-tracking experiment so the same factors apply. Even though the simulation study has a sample of better quality (80+% of frames with eye-tracking data in all trials) but the size of it is actually smaller than the size of the in-field study sample.

On arrivals the identical trends are observed for island stations, where the amended system causes a minor drop in TFD values. As this drop is also combined with increase in cumulative TFC for all elements, it can be suggested that lack of previous experience is an important factor driving cumulative TFD down. As the drivers are developing new search strategies, they potentially switched their attention more often in order to maintain awareness of the updated physical environment. However, it would be incorrect to attribute these trends solely to lack of previous exposure to the proposed layouts as there is no similar trend at Heworth. The results show that spacing of the AOIs can have an important impact on drivers' gaze behaviour. Having the elements located closer in a visual field can bring benefits of reduced saccadic movements thus releasing an additional visual attention capacity.

In scenario 3 cumulative TFD on all AOIs is close to scenario 1 at all stations. Stations with confined layouts have higher cumulative TFD when more passengers are present. This is in line with Chapter 7 results showing more visual attention allocated to platforms in confined environments. As before, it is possible to assume that lack of scenery AOIs leads the drivers to focusing more on the platforms. Furthermore, it shows that more confined station designs can provide metro drivers with a small reserve in visual attention in more demanding situations (high passenger loads). For island stations gazes on the immediate scenery on the left side of the track can be considered as distractions as there are no safety-critical elements. The running signal usually is further ahead. In scenario 3 these distractions remain in combination with higher demands imposed by more passengers. Hence the drivers tend to use shorter fixations and transfer the gaze between the AOIs more frequently.

Results at Heworth also demonstrate benefits of putting the safety-critical elements close together at a narrow FOW angle for departures. Other stations show a drop in cumulative TFD during departures in scenario 2 but it is more pronounced when a distance between the AOIs is increased (reducing the FOW angle between those) than when the FOW angle is increased (reducing the distance).

Despite the changes in cumulative TFD on all AOIs between the scenarios, magnitudes of change between the individual AOIs are different confirming inter-dependence of some elements. This inter-dependence was expected from the onset of this thesis. A necessity to look on one element longer is expected to compromise drivers' ability to devote the same amount as before to other elements. However, it is drivers' visual strategies, perception of risks and motivational pressures, which define what elements will be compromised.

### *8.5.5 Influence of design amendments during arrivals*

In the baseline scenario the drivers prioritised mirrors and stopping position markers on arrivals. Similarly to the real-life driving the participants continue to use mirrors to determine a stopping position. This behaviour confirms that alteration of the monitoring task does not change semantic value of DOO mirrors.

All of the stations demonstrate increase in TFD accompanied by increase in both TFC and AFD values. As mentioned in Section 8.3.5, increase in all three metrics can be a sign of lower situational awareness. This is to be expected from physical environments with amended layouts. Even though that it is also indicative of complexity of information extraction, the trends in the signal-related variables suggest that additional workload is manageable (Holmqvist *et al.*, 2011). The trends for signal fixations also show a possible decrease in task demand compared to the real-life task, where many drivers potentially did not have enough visual capacity to devote to this AOI (Section 7.4.1). It is possible to claim that these trends are not directly related to specific design changes, as a similar situation is observed in scenario 3. The participants stated that they were able to identify the introduced changes. As the signals are the most changed AOI in scenario 2, it is possible that this motivates drivers' to check those more in scenario 3 too. Moreover, indirect effects of design changes can be still present as difference in scan paths can lead to a situation demonstrated in studies on Markov models (Holmqvist *et al.*, 2011). It was found that long fixations in previous AOI would often lead to long fixations on the next AOI (Hooge *et al.*, 2007; Tatler and Vincent, 2008) but it should not affect TFC metrics.

Increased attention to Pelaw signal along with spacing the elements further away from each other leads to longer and more frequent transitions between AOIs. This clearly creates a strain on driver's visual attention and additional workload when monitoring these elements. On the other hand, decrease in the angle between the signal and the mirror does not bring expected performance benefits. The results indicate that reducing the FOW angle only brings workload-related benefits for mirrors and platform indicators when this is accompanied by change in a distance to a running signal. This is shown by Heworth and Gateshead Stadium, whereas performance issues are shown by Felling where the angle was increased. One can note that the direction of change in signal’s position is irrelevant as long as it causes increased visual attention. Unfortunately, the results do not provide an answer on ideal signal position for increased attention as all of the proposed layouts are novel for the participants and incentivise scan path changes.

There are inter-dependencies between the stopping position elements (mirrors and stopping position indicators) and platforms at island platforms. Reverse proportionality can be observed between these elements. Whenever the usability of the stopping position AOIs improves, the participants start to demonstrate increased stress interacting with the platforms and vice versa. While not being able to simulate environmental factors, the effects of the distance to the mirror were checked at Pelaw. Increasing this distance even further (to 13.5m) causes higher workload. Moreover, the drivers were observed stopping past the platform indicator in scenario 2, demonstrating a loss of informativeness of this element. Similar trends to the real-life stopping position indicator at Gateshead are observed with increased stress and decrease in TFD. However, this does not cause any positive effect on the mirror due to novelty of a situation. Moreover, the drivers still stopped quite far away from it.

The participants potentially have a range of comfortable viewing distances to a mirror, which they do not want to compromise. They are willing to sacrifice the platform indication aide if need be. However, this comfortable distance seems to differ between the open and built-over stations, being shorter at the latter locations. It is possible that lighting conditions and narrower platforms are the reason for that. At open stations with wider platforms it is necessary to look on a mirror at a different angle to get situational awareness about as much width of a platform as possible. Whereas at confined stations it is more beneficial to get closer to the mirror to get “zoomed in” view as the wider view does not provide any advantages.

### 8.5.6 Influence of design amendments during departures

The departure results demonstrate same trends as arrival results for running signals, where increase is observed for all three metrics in each location. This supports a notion about poor situational awareness in the amended physical environment. Additional attention and gaze time needed to locate and extract information from a signal in such environment creates usability issues for the second AOI in the departure sequence. For all locations but Gateshead stadium decrease in mirror TFD is rather small and statistically negligible. One can note that spacing between the two elements is decreased everywhere but at Gateshead Stadium. Once again, this indicates that spacing of the elements is an important PSF. However, it is the unusual layout that creates mirror usability issues as this trend can be observed at all stations. It is possible that in long-term the layouts (when drivers are more familiar with those) allowing to retain same levels of TFD on mirrors might actually produce smaller workloads.

### *8.5.7 Influence of increased passenger levels*

For all stations but Gateshead Stadium a notable rise can be noted in total fixation time on platforms. Such outcome of increased passenger levels has been expected and thus supports simulator validity and adequate motivational pressure claims. The results allow studying the drivers' response in terms of their scanning patterns to accommodate for increase in passenger levels.

Most of the island stations demonstrate decrease in TFD on a stopping position indicator in scenario 3. Based on the TFD statistics, the indicator normally is used as a secondary tool when selecting a stopping position. Hence the participants might decide to re-allocate some of the fixations from it to platforms when there is a need. Moreover, in busy environments the marker can become obscured by passengers standing close to a platform edge, especially from longer distances. As the chosen eye-tracking analysis method uses static AOIs for dynamic environment, it is possible that some of the fixations on one of the nearby AOIs can be recorded as fixations on another AOI. This is what potentially is causing only a minor TFD increase on Gateshead Stadium platform. Similarly, passengers on a busy platform at Heworth can substantially obscure a mirror and lead to erroneous fixations being recorded. When a mirror is actually obscured, the participants switch their attention to a stopping position indicator without no stress-related trends, as it happened at Heworth and in the real-life eye-tracking experiment at the confined stations. In a sense, this is an example of several defences in a system allowing the drivers to stop in a correct position. The participants show that they can easily switch from one to another.

All mirrors on the island platforms demonstrate increase in workload and associated stress. This happens in conjunction with considerable drop in mirror TFD at most of these locations. Being the primary element for choosing a stopping position any distraction from it have a potential to cause increased workload. The drivers confirm that they believe the passengers create distractions (Section 6.3.1) and it is known for distractions to increase workload and stress of a main task (Burns *et al.*, 2002; Pitsopoulos *et al.*, 2010). Work-related distractions were found to be a contributory factor in majority of rail incidents by Madigan *et al.* (2016). RSSB *et al.* (2006) claim that there is a strong correlation between passenger levels and number of incidents. This is further supported by the historic incident analysis where many incidents correlate with passenger levels. Such division of drivers’ attention can cause additional stress as there is not much additional capacity available. According to Naweed and Balakrishnan (2014), operational environment becomes a major factor influencing urban rail drivers’ visual and driving strategies.

### 8.5.8 Failure to check a signal aspect on departure

The failures to check a signal are still recorded even though their rate decreased significantly compared to the real-life study. Assuming the same motivational pressures for the participants, one can note that the considerably better sample quality leads to this. This confirms that some of the violations disregarded in Section 7.4.5 are indeed cases of the eye-tracking set not recording a fixation. Nevertheless, it also shows that the remaining failures could have been genuine near misses. This is a serious alarm for Tyne & Wear Metro as such events are precursors to SaSSPaDs. However, there is a possibility that other factors described in the in-field study, e.g. side gazes, are present in the simulator experiment. As the eye-tracking equipment is the same for both experiments there are no additional control mechanisms against those factors.

Assuming that all of the cases are genuine violations, one can note that scenarios with the baseline layout have the highest rate of such events. Pelaw and Felling are the locations where the most incidents happen. These results are similar to the findings of the in-field eye-tracking study, confirming that the same factors apply in a more controlled environment. Among the design PSF are a distance to a signal and the FOW angle. Spacing of the elements is not as important when it comes to the signal checking violations.

Pelaw and Gateshead Stadium are the only stations in scenario 2, where the distance between a stopping point and a signal is more than 30 metres. The distance at Pelaw is 3 metres higher. Despite the narrower FOW angle, the additional distance results in more failure events at Pelaw than at Gateshead Stadium. On the other hand, the FOW angle at Pelaw decreases in scenario 2 compared to other scenarios. This results in 50% drop in failures to check the signal. The FOW angle increases at

Felling but the distance is reduced considerably in scenario 2 so no violations are observed. It is possible to conclude that decreasing the distance from a stopping point to a running signal below 30 metres can have significant positive effect on the drivers' signal checking behaviour. If such a change is impossible or impractical, reducing the FOW angle might provide an additional incentive to check the signal.

## 8.6 Conclusions

The simulation experiment has shown that the use of modern technologies allows creating low budget high fidelity simulators. These simulators can re-create the immediate physical environment in the VR and provide the practitioners with an opportunity to alter scenarios on the go. Moreover, this kind of simulators is easier to upgrade given the rapid pace of technical development in the field.

The novel approach in validation of such simulators, which is used in this experiment, also proved to be successful. Using the AOIs in dynamic environment reinforces validity results. Previous approaches using static AOIs in the immediate physical environment do not provide full picture of drivers' behaviour and performance due to a strong link to between performance and environment outside of a cab. Validity of the developed simulator can be primarily improved by using a control device that is similar to a metrocar master controller, e.g. a flight simulator joystick. However, these controllers have to be first supported by the software. This highlights one of the disadvantages of the disadvantages of the explored approach as the end-users do not have full control over the software capabilities.

Many assumptions made after the in-field eye-tracking experiment have been supported in a more controlled environment. Drivers behave differently in more confined station environments than at open stations. For example, the interaction with mirrors and a comfortable distance threshold depends on the station type. Lack of scenery-related distractions keeps drivers' focus on elements of station design in more confined locations. The mirrors are used as a primarily mean to select a stopping position. Stopping position indicators or platforms are the second most fixated at AOI depending on passenger loadings. There is a clear inter-connection between platforms and the AOIs used to select stopping position.

The drivers tend to focus on the signals more in the amended system even when no changes are introduced to signals design. It is possible that increased focus on SPaDs in the industry forces the participants to check a signal for changes first when they expect amended layouts. It is also possible that task demand in simulator is lower subsequently freeing up some visual attention for signals. Even though hypothesis 8a is not supported by the test results, visual attention on the signals is inter-dependent with the effects introduced by change in the FOW angle. The workload-related benefits of the reduced FOW angle on arrivals can only be observed in conjunction with an increased focus on a signal. Hence the hypothesis 8b is supported for arrivals but subject to other factors. Hypothesis 8b is not supported by the departure results, where an increased attention to one of the elements causes the usability issues for the second element regardless of the layout. On the other hand, the closer spacing of a signal and a mirror is found to be a stress-reducing PSF.

The spacing is not as important as the FOW angle and the distance to a signal when it comes to the violations to check a signal before departure. The departure statistics support hypothesis 8c. The combination of a signal located closer than 30m away from a stopping point and a narrow FOW angle proved itself as the best layout to improve the signal checking behaviour.

Drivers' ease of interaction with the stopping position elements suffers when passenger levels increase thus supporting hypothesis 8d. More commuters on the platforms create a distraction from a stopping task resulting in gaze performance that indicated increased workload and stress. However, the results also show that the participants are able to switch from a mirror to a stopping position marker for choosing a correct stopping point in case the mirror becomes obscured or unusable. This switch happens without any additional stresses associated with the markers. Hence the designers need to make sure that a secondary stopping position aid is present at a station that is informative. Hypothesis 8e is not corroborated as only a small number of variables is dependent on drivers' experience. However, overall experience with the system (regardless of driving experience) might play an important role in driving. Chapters 5-7, however, show that even unexperienced drivers have a good knowledge of the existing system.

# Chapter 9. Final conclusions

## 9.1 Thesis conclusions

The thesis presents a comprehensive assessment of Tyne & Wear Metro drivers' performance and effects of system design PSFs on it. The assessment approaches the research problem from different positions, taking into account different contexts affecting metro drivers' tasks. Using Tyne & Wear Metro as a case study allows multiple comparisons with the past research due to a variety of design features across the system. Metro-specific research has significant potential for bringing safety benefits to such systems. Many of the findings of this thesis are associated with design features, which more often feature in urban rail, e.g. DOO, high passenger loadings, shorter headways and simpler signalling. Hence there is a high transfer potential to other systems for some findings as long as necessary attention is given to the spatial context in the process.

The assumptions made at the onset of the research about metro systems being rather unique have been validated and Research Question 1 answered. Many concepts and findings from mainline railways or other safety-critical industries are not applicable or do not act in the same way in urban railways, particularly in system design. In context of Research Question 2, Tyne & Wear Metro as the DOO system implies different priorities among the drivers, where station dispatch related pressures can be more important than sticking to green signals. One of the unexpected findings is related to route knowledge, which is definitely very high. However, in certain situations it acts as an adverse PSF due to drivers not being ready to non-routine operations. Even though Tyne & Wear Metro drivers do not suffer from the task monotony (partially due to DOO), the system itself is monotonous and highly repetitive. This allows for quick learning but does not prepare for disruptions. When combined with lack of advance warning and decreased situational awareness, it becomes one of the main reasons of incidents.

The research uncovered potentially hazardous behaviour by metro drivers not checking the signals before departure from a station. Such violations are precursors in SASSPaDs. Even though there are concerns about the equipment used affecting these results, these events have to be taken seriously. On the other hand, it shows yet another discrepancy with non-DOO systems where the drivers have less concurrent tasks. Metro drivers tend to prioritise DOO equipment monitoring over the signal checks, potentially due to high route knowledge and the PTI related hazards.

The results support theories about importance of exploring everyday behaviour and what is considered “normal operations”. Many features have been part of Tyne & Wear Metro for a long time hence might be considered safe reactively. For example, having a signal near the platform edge is a normal practice. However, it was shown that such alignment of track circuits might induce unsafe behaviour. It is unknown whether moving the signal further from the platform would create any additional risks but it would address the divided attention issue on departure. In fact, the stations without running signals in Tyne & Wear Metro, e.g. Cullercoats, have the least number of any driver-related incidents.

Most of the design related PSFs discovered at the latter stages of this research are adverse due to the issue of divided attention. It is possible that distractions highlighted by the drivers are often a cause of this issue too. At busy stations the decrease in visual interaction with mirrors and stopping position markers was observed.

Research Question 3 is successfully addressed in Chapter 8. Station layouts should be able to mitigate the issue of divided attention by supporting smaller saccadic movements, providing fit for purpose stopping position aide or simply prioritising one if the platforms are busy. Among the design interventions are reduction of the FOW angle, reduction of the distance to a running signal to less than 30m, closer spacing of a mirror and a signal on a longitudinal axis. The first two interventions should also improve signal checking behaviour. However, the interventions should be based on a local assessment of operational risks, and other PSFs such as openness of the stations, lighting conditions, potential for disruptions etc.

System design features, according to the drivers, do not have a significant effect on their performance even though the drivers can differentiate between those in terms of usability. More discrepancies were found between the drivers’ opinions and objective measures but some can be attributed to more than 6 years long period under investigation throughout a number of phases. The historic incident analysis shows how incident statistics can drastically change in less than a year due to the operator’s initiatives.

The research in railway industry still underutilises modern technologies, e.g. eye-tracking, compared to other safety-critical domains. Even though the safety concerns exist about experiments in operational environment, the thesis showed that such studies are possible. The new methodologies allow producing rich data samples through which train drivers' cognitive processes and performance can be studied. In long-term these technologies should help establishing objective baseline values for human performance, which then can be used in assessments of operators and infrastructure. One of the achievements of this research is a demonstration of a methodology to investigate gaze behaviour on small dynamic elements outside of the cab environment.

The assessments, if safety concerns exist, can be conducted in simulated environments. The current technological advancements make the PSF research a much more affordable undertaking. The thesis shows that nowadays it is possible to build a high fidelity simulator with a budget times smaller than required for conventional setups. The user-friendly tools provide safety staff with means of re-creating specific scenarios, which can be then used for training and assessment. Moreover, the combination of the two methods used in this research (eye-tracking and simulators) proved to be a good approach for validation of each other, which is an important pre-requisite for an industrial implementation.

## 9.2 Main contributions

The author believes that this thesis provides a number of contributions to the existing body of knowledge. This section will elaborate on these.

### *9.2.1 Transferability of mainline knowledge to metro-specific PSF research.*

The literature review demonstrates that PSF research in metro systems is scarce and urban railways are often considered to be the same as mainline railways. The following experiments and analysis prove that many of the PSFs described in the body of knowledge act differently or are affected by unique factor combinations in urban railways. This misalignment had gone almost unnoticed amongst researchers until now.

As the field of rail human factors developed, it has been acknowledged that transferability of knowledge from other industries might be limited due to different operational contexts (Kyriakidis *et al.*, 2015). However, there might be a need to introduce a further division within the railway human factors field to address differences in rail systems of various types. This thesis uses the urban railway, namely Tyne & Wear Metro, to show that transferability from mainline railways is rather limited.

The growth in the number of urban rail systems around the world creates additional pressures to understand metro-specific incident propagation mechanisms and mitigation options. As with mainline railways, PSFs can have an important role in this process. However, the current body of knowledge in rail PSFs mostly consists of the work focusing on mainline railways. It is often retrospective and generalised to achieve a better fit to a wider number of systems. This brings benefits of using bigger samples but the complexity of interactions of PSFs in each unique system is somewhat overlooked. For example, organisations like RSSB still group metro incidents with mainline incidents for the research purposes. Furthermore, the body of knowledge is also skewed towards certain incident types which may not be as risk-bearing in urban rail systems where speeds are lower but station stops are significantly more frequent.

There has been a move towards recognising the need for in-depth PSF research focused on urban railways recently (Kyriakidis *et al.*, 2012a; Naweed and Balakrishnan, 2014) but the work in this area is still rather fragmented. What is more important, it is that it is still unknown whether findings from mainline railways are applicable in urban rail and vice versa. Moors *et al.* (2015) claimed that such transferability should consider local characteristics of a system. Hence research grouping data from multiple railways does not take into account a holistic notion of a system as presented by Dekker (2001) where the environment, tools and tasks are all important factors in incident propagation.

The thesis demonstrated that findings from the mainline railways are not always applicable to Tyne & Wear Metro and thus need to be treated with caution for urban rail. The already mentioned focus on SPaD incidents does not reflect the risk profiles seen in Tyne & Wear Metro where almost 50% of the incidents are related to station procedures.

It has been shown that DOO drivers can have different motivational pressures to those of mainline train drivers. Even though it has been acknowledged previously that the risks of the SaSSPaDs are higher in DOO systems (Basacik *et al.*, 2009; Naweed and Rainbird, 2014; Multer *et al.*, 2015), there was no work done to explore what it means to train driving tasks in such systems. Furthermore, it has been claimed that drivers are at a higher risk of not checking the PTI due to time pressures (Basacik *et al.*, 2009). The thesis shows that Tyne & Wear Metro drivers clearly prioritise the PTI-related tasks over signal checking thus increasing risks of SPaDs, not the PTI-incidents.

Another aspect of the non-transferability found is operation in tunnels, which is one of the main features of many urban rail systems. There is very little evidence in the literature on how it affects train drivers. Yang *et al.* (2012) demonstrated that drivers' arousal levels increase as luminance increases. Arousal levels however drop in monotonous environments. Tunnels are universally accepted as monotonous environments but the analysis in Chapter 5 did not support these assumptions. It is known that station stops cause disengagement from a driving task (Naweed and Rainbird, 2014) and increase arousal levels (Yang *et al.*, 2012). Hence short distances between stations (a key characteristic of urban rail) lead to a situation when the adverse effects of monotony in tunnel sections might be balanced out by other events, e.g. station stops.

Although it is known for non-routine events to be the contributory factors in incident propagation (Madigan *et al.*, 2016), there is no research to date focusing on the relationship between repetitiveness of metro train driving and situational awareness in context of such events. Route knowledge is often mentioned as a positive PSF in train driving due to its beneficial effects on signal sighting (Naweed, 2013; Gibson, 2016). The results of this thesis support the initial assumptions of metro drivers having very high route knowledge due to operating in a simpler closed system. This is a result of the high repetitiveness of a task, and comes at a cost of formed expectation biases and lower situational awareness (O'Connell *et al.*, 2015). Therefore drivers are more prone to making errors in the non-routine events, as demonstrated by this thesis. The findings show that high route knowledge, contrary to the assumptions found in the literature related to mainline applications, can be an adverse PSF in urban rail.

Findings from mainline railways indicated significant effect of personal factors, namely experience, on visual strategies (Li, 2004; RSSB, 2005). Based on the review of findings in other industries it was expected that experienced drivers would have more sophisticated and efficient strategies. The experiments conducted in this thesis demonstrated only limited effects of previous experience, especially in a familiar system. What is more important, the failures to check a signal were found in the in-field eye-tracking study where only the experienced drivers participated. This suggests that there might be an inter-dependency between the system design (DOO) and the personal factors (experience) which induces propagation of these violations.

### *9.2.2 Evidence based identification of system design PSFs and mitigation measures.*

The conducted experiments demonstrate importance of a system design in influencing metro drivers' behaviour. Selected elements of a station design have shown their ability to affect drivers in a series of objective tests. Importance of exploring PSFs in the infrastructure designs that are used currently is highlighted by the fact that all the elements demonstrating adverse effects on drivers' performance currently comply with Railway Group Standards. A number of design interventions has been proposed and validated experimentally in order to mitigate these adverse effects.

The previous research showed that physical environment and the associated factors are important elements shaping front line staff performance. The detrimental effects of system designs were demonstrated in many incident investigations. However, the extent of such effects and effective mitigation strategies are still largely unknown with studies being fragmented.

System design factors, although featuring in all reviewed taxonomies, are often overlooked in rail research. As the current body of knowledge is skewed towards SPaD incidents, many system design factors explored relate to signal sighting. Furthermore, task or organisational factors can be similar across a range of industries but it is rarely a case for system design factors. Hence it is harder to transfer knowledge from other industries compared to knowledge on, for example, the safety culture or fatigue management. The findings from mainline railways which were corroborated in this thesis are also related to the task factors, e.g. decrease in arousal levels after 2 hours of driving. It is also harder to conduct meaningful experiments to explore system design PSFs as those require good access to infrastructure and are associated with many constraints, as discussed in Section 9.2.4.

The thesis explored the system design factors which are common for urban rail systems. For example, a confined nature of stations, sharp drops in speed limits and a simpler signalling were found to be associated with poorer drivers' performance. At the same time, the Type 1 design station and simplified station layouts demonstrated positive effects on the incident statistics.

An in-depth investigation of the station layouts standard for Tyne & Wear Metro allowed identifying the PSFs associated with individual elements, e.g. DOO mirrors or stopping position indicators. Using the non-intrusive techniques helped explore the station layouts from a day-to-day operations perspective and uncover important relationships. For example, a spacing of a signal and a mirror has been found to reduce workload. On the other hand, a FOW angle and a distance to a signal were found to affect the signal checking performance on departures. These findings also supported a notion of some layouts being worse for drivers' performance. However, the lack of previous research in this area allows these layouts to be a standard practice in urban rail systems thus exposing drivers and passengers to an unnecessary risk.

The thesis proposes a set of design improvements which should help create an optimum layout for systems like Tyne & Wear Metro. This is even more important as some of the design features typical for Tyne & Wear Metro are considered for implementation in other types of railways. For example, DOO driving is currently becoming a big focus area in mainline railways to cut staff costs. Hence better understanding of drivers' train dispatch routines can inform decision making and implementation of DOO elsewhere. Moreover, it is easier to implement the proposed design mitigations while a system is being converted to new operational practices.

The findings of the thesis can be transferred to other railways, especially if those use a similar system design. Key areas which need to be compared to Tyne & Wear Metro to transfer the findings are the dispatch procedures, the signalling design and the station layouts. Train operating companies should start focusing on station procedures rather than signal sighting as a large portion of risk to passengers sits there. This work clearly demonstrates these risks. The notion of inter-dependency of PSFs is reinforced by several experimentally proven examples which can be applied in other systems where operational conditions are similar to Tyne & Wear Metro.

### *9.2.3 Inter-dependencies of system design PSFs uncovered.*

The thesis has been built around the idea of exploring driver-related incidents from a system point of view when a complicated web of factors affecting front line staff needs to be considered. The experimental results clearly indicate that drivers' interaction with the physical environment changes along with environment and operational conditions. The work done provides insights into the inter-dependencies of PSFs thus highlighting the importance of this largely unexplored area of human factors.

Every system contains a multitude of factors related to a design, a task and operational environment. The factors are connected dynamically and the links can change depending on the state of these. A human operator is there to create safety in a system while interacting with a variety of factors simultaneously (Dekker, 2001). Therefore, researching just one factor will not provide a full picture as it can have positive effects in specific circumstances but become a negative PSF when surrounding conditions change. By understanding the dynamics of how various factors affect PSFs, it is possible to build more in-depth models of train drivers' performance. These models would exist for each operational scenario in a system and inform system engineers on potential improvements or mitigations.

The research that acknowledges inter-dependency of PSFs is beginning to appear although some links, especially with a system design, are missing. Even though environment and a design are the key PSF producing areas (Dekker, 2001), the previous work on the inter-dependence of factors in mainline railways (Kyriakidis, 2013; Kyriakidis *et al.*, 2015) does not account for the connections between the two areas. There is very little work available on the inter-dependencies between system design PSFs too. However, the notion of inter-connected factors implies that altering one element of a system design should affect factors associated with other elements.

In a sense, the whole eye-tracking methodology in this thesis revolves around inter-dependency of PSFs because a connection between a system design and a task (workload and stress) is assessed. This is one of the few well-known inter-dependencies and has been explored in many studies previously (Basacik *et al.*, 2009; Blanchard and Lowel, 2009; Johnson *et al.*, 2009; Hitchcock *et al.*, 2013; Kyriakidis *et al.*, 2015). However, this connection is rarely considered within multi-factor (more than 2) relationships. Going back to the notion of a web of factors in a system constantly interacting between each other and dynamically changing, it would be incorrect to assume that the 2-factor relationships can exist in isolation.

The thesis demonstrated the multi-factor inter-connectivity in the assessment of the system design PSFs in the series of experiments. Many relationships found include the task PSFs (workload, stress, visual strategies) and at least 2 system design PSFs. For example, the station "openness" is the PSF affecting the signal layout –

workload PSF relationship. The usability and ease of interaction with stopping aids significantly depends on design of mirrors and passenger approaches. Similarly, drivers' priorities in design elements are dependent on these factors.

The experiments showed that the drivers' interaction with all tested elements are inter-dependent, where some PSFs can act as catalysts in other relationships. Existence of such PSFs has been suggested in the beginning (Section 5.4.7) by demonstrating that a number of trains calling at a station can induce adverse effects of other factors. This has been proven later in the experimental phase. For example, a signal attracting more visual attention enhances the positive effects of the infrastructure layout with narrower FOW angle. The narrower FOW angle can enhance signal checking behaviour but the benefits of this PSF can be improved if it is combined with a shorter distance between a stop point and a signal.

The multi-factor relationships have also been experimentally explored with several PSFs not related to a system design. Passenger levels at stations were found to interact with system design and task PSFs, where the system design PSFs act differently with various passenger loadings. An interaction with stopping elements on arrivals suffers (workload increases) due to the need for more monitoring of platforms. Drivers' visual strategies change when there are more passengers on platforms due to prioritising stopping elements with better visibility. It is also possible to argue that the design of platforms and passenger approaches affects how platforms are perceived by drivers, e.g. crowded or not, thus becoming another PSF in the multi-factor relationship described above.

The personal factors (experience) became another PSF involved in the inter-dependencies although this involvement is rather limited. With drivers utilising different visual strategies based on their experience levels, it is possible to see how the connections between the task and system design factors can be affected.

The thesis demonstrated the presence of the inter-dependencies between the PSFs and the complex nature of factors affecting drivers in urban railways. Although experimentally only a few relationships are shown there is a much wider potential for research in this area. Understanding of PSFs outside of the isolated 2-factor relationships is a key to the assessment of drivers' performance from a system perspective. Beneficial and adverse effects of some PSFs were proven experimentally thus supporting the need for the industry to explore those.

### *9.2.4 Validated low-cost methodology for system design PSF research in railways.*

The thesis proposes a methodology which allows non-intrusive research in railway design PSFs using virtual reality environments. User-friendly software, with a simple setup and a low cost makes this methodology easily adaptable for various research projects. Moreover, the thesis validated the proposed method through the comparison of a train drivers' performance in a simulator and real-life driving with the 90% match achieved.

An experimental approach into ergonomics of a system usually requires comparisons between several options. In the case of railway infrastructure, it can be quite intrusive and there is a necessity for an informed executive decision to be made before changes are introduced into a system. For example, assessing potential safety benefits of introducing a new signal type requires introducing risks associated with a number of issues, e.g. system closures for installation, driver training, rule book changes. Hence it must be a well-tested signalling product to mitigate against such risks. Nowadays this means a series of high fidelity experiments, using test tracks or expensive virtual reality setups. This, in turn, leads to significant costs for the introduction of new solutions in the industry. According to Simmons (2015), the "safety first" approach and budget constraints are two main barriers for the introduction of new technologies and solutions in the railway industry.

The research in design-related PSFs is constrained by the same barriers as it often requires a similar comparative experimental approach. The thesis tries to address the monetary constraint by proposing a methodology for a low cost high fidelity virtual environment which can be developed and altered quickly. Moreover, it is a setup that utilises only widely available off-the-shelf products. Until recently it was acknowledged that the high fidelity simulators always come with an inflated price tag due to the need for a physical shell, moving base etc (Jamson and Jamson, 2010; Jolly *et al.*, 2013).

The literature suggested that the video games industry will be able to contribute to the field of railway simulation (Rushby and Seabrook, 2007). The low cost alternatives to the full simulator setups have been explored previously (Yates and Sharples, 2009; Reed and Green, 2010; Bella, 2014) and even rolled out commercially (Rowe, 2013). However, the research in system design rail PSFs has not adopted this approach yet and the expensive setups continue to be used (Aitchinson and Davies, 2009; Buksh *et al.*, 2013).

The thesis shows that the cheaper, more flexible and adaptable approaches can be also successful, validating this in the series of experiments. Many rail human factors studies use subjective measures due to financial limitations whereas the proposed methodology allows collecting the rich objective data. The thesis demonstrated that the subjective measures, e.g. questionnaire results, do not always reflect drivers' actions out in the field. Hence collecting the objective data improves the reliability of the research.

As with many other technologies, a decrease in cost usually causes an increase in uptake of a certain technology. Therefore, the proposed methodology has the potential to boost research in system design rail PSFs. Furthermore, significant benefits have been identified for train operators who would like to start using virtual reality solutions in their training and assessment process to drive down the costs of training. There is a requirement for a minimum number of hours driven by a trainee driver before he or she can become fully qualified. These hours come at a significant cost to a train operating company. The proposed methodology has a high cost-benefit ratio addressing this issue.

## 9.3 Limitations

Although a set of contributions to the body of knowledge has been delivered, there are limitations to this research.

### *9.3.1 Tyne & Wear Metro operational context*

A potential limitation of the findings of this thesis is related to the uniqueness of urban rail systems which strongly depend on particular characteristics, e.g. topography, patronage. These characteristics might differ across systems. Nevertheless, Tyne & Wear Metro is a good example of urban rail systems as it contains features common to various types of systems as discussed in Chapter 4 (Table 6). This makes it a suitable subject for a case study to explore the design-related PSFs.

### *9.3.2 Technical limitations of the eye-tracker used*

The eye-tracker used did not allow for a meaningful investigation of driving in tunnels which is one of the main design features of metro systems. Although Chapter 5 and Chapter 6 show no links between driving in tunnels and increase in incidents due to monotony, it still can be expected that drivers' visual strategies might be different in underground environments (as partially shown by the confined stations Heworth and

Gateshead Stadium in Chapter 8). Hence the design interventions proposed might not bring the assumed benefits in such environments.

There are some limitations on the eye-tracking methodology used. Use of physical infrared tags can help increase a sample size quickly. The tags automatically register fixations on each AOI thus removing any ambiguities created by a researcher specifying the AOIs manually and the 3-second timeframes. For this thesis the required number of tags (16) was not available as those have been discontinued for Tobii Glasses (Mark 1). More modern devices will have these add-ons widely available. However, it is still necessary to check the off-sets and (if needed) repair the sample manually as described in Section 7.2.6.

The accuracy of measurements can also be increased by using eye-trackers with a better sampling frequency. As shown by Anderson et al (2010), lower frequency increases the probability of a measurement error. There are eye-trackers available on the market with the sampling frequencies in excess of 300Hz. This parameter can also contribute to an improved data quality and reduce exclusion of trials from a sample.

### *9.3.3 Metrics used to investigate changes in workload/stress*

Chapter 7.2.4 presents a strong case for a link between the fixation count, fixation duration and stress/workload. Many eye-tracking studies focusing on workload and stress use measures of blinks and pupil size (Marquart *et al.*, 2015). For example, Schulz *et al.* (2011) concluded, in their comparison of the eye-tracking metrics, that the pupil diameter metrics might be the best measure to estimate the workload.

These metrics are not collected by the equipment used in this thesis but are available in other similar devices. It might be beneficial to use the metrics associated with blinks and pupil size in similar studies in the future to reduce the ambiguities introduced by comparing the numerous fixation metrics.

### *9.3.4 Limitations of the simulator used*

The simulator provides good visual stimuli but the simplifications were introduced in terms of the fidelity of the train controls. There are no controller devices available which resemble a Tyne & Wear Metro train drivers' desk. However, the results obtained are still considered significant with the expected variations in performance with different controls being small.

## 9.4 Recommendations for Further work

Future research should explore effects of the design amendments in the high passenger flow situations as the reverse relationship is present between the platforms and stopping position elements. The future experiments also need to improve drivers’ familiarity with the amended layouts before the tests are conducted in order to contain effects of poor situational awareness. Using the more advanced eye-trackers would allow investigation of the PSFs present in metro tunnels and at underground stations.

As the thesis progressed through the phases, many promising research areas have been taken out of the scope. Effectively, the further research should explore in depth the potential of each PSF identified in this thesis, i.e. the effects of metro environment on arousal levels, the operations in tunnels in terms of ambient lighting and DOO monitors, the presence of disruptions and the areas with ground position lights, and other. It is even more important to conduct this research in urban rail as the body of knowledge in this domain is comparatively small.

An alignment of several PSFs is an important pre-condition in the incident causation. Although the thesis uncovers some of these relationships at a design elements level, the full dependence web is yet to be understood. This research domain promises significant gains in understanding of the causal mechanisms and human performance.

# References

Afrooz, A. E., White, D. and Neuman, M. (2014) 'Which Visual Cues Are Important in Way-Finding? Measuring The Influence Of Travel Mode On Visual Memory For Built Environments', in Caltenco, H., Hedvall, P. O., Larsson, A., Rassmus-Grohn, K. and Rydeman, B. (eds.) *Universal Design 2014: Three Days of Creativity and Diversity*. Amsterdam: IOS Press BV.

Ahmadian, M. and Vahdati, N. (2006) *Rail Conference, 2006. Proceedings of the 2006 IEEE/ASME Joint*. 4-6 April 2006.

Aitchinson, J. and Davies, K. (2009) *Drivers' response to junction signalling: A simulator study (T669 Report)* London.

Andersson, R., Nyström, M. and Holmqvist, K. (2010) 'Sampling frequency and eye-tracking measures: how speed affects durations, latencies, and more', *2010*, 3(3).

Andrews, T. J. and Coppola, D. M. (1999) 'Idiosyncratic Characteristics of Saccadic Eye Movements When Viewing Different Visual Environments', *Vision Research*, 39(17), pp. 2947-2953.

Association of Train Operating Companies (2006) *ATOC/GN009: ATOC Guidance Note - Investigation of Station Overruns and Failures to Call*. London: ATOC.

Australian Transport Safety Bureau (2015) *Incident Involving Absolute Signal Blocking* Canberra: ATSB.

Balfe, N., Sharples, S. and Wilson, J. R. (2015) 'Impact of automation: Measurement of performance, workload and behaviour in a complex control environment', *Applied Ergonomics*, 47, pp. 52-64.

Basacik, D. and Gibson, H. (2015) 'Where is the platform? Wrong side door release at train stations', in *Contemporary Ergonomics and Human Factors 2015*. Taylor & Francis, pp. 441-448.

Basacik, D., Read, C., Heavisides, J., Jones, M. and Pollard, G. (2009) *A review of passenger train dispatch from stations (T743 Report)*. London: RSSB.

Basacik, D., Waters, S., Reed, N. and Whitmore, A. (2015) *A pilot study of low workload in train drivers*. London: RSSB/TRL.

Baysari, M. T., Caponecchia, C., McIntosh, A. S. and Wilson, J. R. (2009) 'Classification of errors contributing to rail incidents and accidents: A comparison of two human error identification techniques', *Safety Science*, 47(7), pp. 948-957.

Baysari, M. T., McIntosh, A. S. and Wilson, J. R. (2008) 'Analysis of Australian and UK rail incidents and accidents using Retrospective TRACEr lite for train drivers', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 52(21), pp. 1761-1765.

Beanland, V. and Pammer, K. (2010) 'Looking without seeing or seeing without looking? Eye movements in sustained inattentional blindness', *Vision Research*, 50(10), pp. 977-988.

Bearfield, G., Holloway, A. and Marsh, W. (2013) 'Change and safety: decision-making from data', *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 227(6), pp. 704-714.

Beck, A. T. (1985) *Anxiety disorders and phobias : a cognitive perspective*. New York: New York : Basic Books.

Belamy, L. J. and Geyer, T. A. W. (2007) *Development of a working model of how human factors, safety management systems and wider organisational issues fit together*. Bootle: HSE.

Bella, F. (2005) 'Validation of a Driving Simulator for Work Zone Design', *Transportation Research Record Journal of the Transportation Research Board*, 1937(1), pp. 136-144.

Bella, F. (2014) 'Driver perception hypothesis: Driving simulator study', *Transportation Research Part F: Traffic Psychology and Behaviour*, 24, pp. 183-196.

Blaikie, N. (2003) *Analyzing Quantitative Data: From Description to Explanation*. London: Sage Publications UK.

Blair, J., Czaja, R. F. and Blair, E. A. (2013) *Designing surveys: A guide to decisions and procedures*. SAGE Publications, Incorporated.

Blanchard, H. (2013) 'A method for assessing drivability for ETCS cab retrofit', *4th International Conference on Rail Human Factors* London. CRC Press, pp. 191-200.

Blanchard, H. and Lowel, E. (2009) 'Using a model of train driver performance to investigate the contribution of human factors to rail incidents ', *3rd International Conference on Rail Human Factors*. Lille.

Bogner (2002) 'Stretching the Search for the 'Why' of Error: The Systems Approach', *Journal of Clinical Engineering*, 27(2), pp. 110-115.

Bonnett, C. F. (2005) *Practical railway engineering*. 2nd ed.. edn. London : Hackensack, NJ: London : Imperial College Press
Hackensack, NJ : Distributed by World Scientific Pub.

Boring, R. L., Griffith, C. D. and Joe, J. C. (2007) *Joint 8th IEEE HFPP Conference on Human Factor and Power Plants*. Monterey.

Borji, A. and Itti, L. (2014) 'Defending Yarbus: Eye Movements Reveal Observer's Task', *Journal of Vision*, 14(29).

Borji, A., Lennartz, A. and Pomplun, M. (2015) 'What do eyes reveal about the mind?: Algorithmic inference of search targets from fixations', *Neurocomputing*, 149, Part B, pp. 788-799.

Borkan, J. M. (2004) 'Mixed Methods Studies: A Foundation for Primary Care Research', *Annals of Family Medicine*, 2(1), pp. 4-6.

Box, G. E. P. and Cox, D. R. (1964) 'An Analysis of Transformations', *Journal of Royal Statistical Society. Series B (Methodological)*, 26(2).

Bremond, R. (2000) 'The Conspicuity of Traffic Signs in Complex Backgrounds', *Color Research & Application*, 25(5).

Brooke, J. (1986) *SUS - A Quick and Dirty Usability Scale*. Reading: Ltd., R. C.

Bryman, A. (2014) *Social Research Methods (4th edition)*. New York: Oxford University Press.

Buksh, A., Sharples, S., Wilson, J. R., Coplestone, A. and Morrisroe, G. (2013) 'A comparative cognitive task analysis of the different forms of driving in the UK rail system', *Fourth International Rail Human Factors Conference*. London. Taylor & Francis group, pp. 173-182.

Burns, P. C., Parkes, A., Burton, S., Smith, R. K. and Burch, D. (2002) *How Dangerous is Driving With a Mobile Phone? Benchmarking the Impairment to Alcohol*. London.

Caird, J. K., Creaser, J. I., Edwards, C. J. and Dewar, R. E. (2002) *A HUMAN FACTORS ANALYSIS OF HIGHWAY-RAILWAY GRADE CROSSING ACCIDENTS IN CANADA*. Ottawa: Canada, T.

Carter, C. J. and Laya, O. (1998) *Drivers' visual search in a field situation and in a driving simulator*.

Castro, C. (2009) *Human Factors of Visual and Cognitive Performance in Driving*. Boca Raton: CRC Press.

CCD Design & Ergonomics Ltd (2005) *Assessing the impact of increased numbers of CCTV images on driver only operations of trains* (T535). London.

Ceci, R., Patten, C. and Mårdh, S. 8020 LNAI (2013) 'An evaluation of the interior design of the Stockholm bypass tunnel - A driving simulator study'. pp. 3-11.

Celik, M. and Cebi, S. (2009) 'Analytical HFACS for investigating human errors in shipping accidents', *Accident Analysis and Prevention*, 41(1), pp. 66-75.

Chang, H. L., Yang, C. Y. and Wu, J. G. (2005) 'Job stress and its influence on train drivers', *Tiedao Xuebao/Journal of the China Railway Society*, 27(2), pp. 21-27.
Cheng, A. S. K., Ng, T. C. K. and Lee, H. C. (2011) 'A comparison of the hazard perception ability of accident-involved and accident-free motorcycle riders', *Accident Analysis & Prevention*, 43(4), pp. 1464-1471.
Choi, M.-H., Min, Y.-K., Kim, H.-S., Kim, J.-H., Yeon, H.-W., Choi, J.-S., Kim, B., Min, B.-C., Park, J.-Y., Jun, J.-H., Yi, J.-H., Tack, G.-R. and Chung, S.-C. (2013) 'Effects of three levels of arousal on 3-back working memory task performance', *Cognitive Neuroscience*, 4(1), pp. 1-6.
Clarke, S. (1998) 'Organizational factors affecting the incident reporting of train drivers', *Work and Stress*, 12(1), pp. 6-16.
Clements, M., Minson, R., Moisio, L., Moll, J., Moore, J. and Scott, A. (2011) *Eco-driving: understanding the approaches, benefits and risks (T839 Report)* London.
Coeckelbergh, T. R. M., Cornelissen, F. W., Brouwer, W. H. and Kooijman, A. C. (2002) 'The Effect of Visual Field Defects on Eye Movements and Practical Fitness to Drive', *Vision Research*, 42(5), pp. 669-677.
Community of European Railway and Infrastructure Companies (2009) *Guidelines for the Psychological Assessment of Train Drivers and other safety related personnel*. Brussels.
Connor, P. (2007) 'The Underground Electric Train', *London Underground Railway Society* [Online]. Available at: http://lurs.org.uk/documents/pdf%2007/march/The%20Underground%20Electric%20Train%20march.pdf (Accessed: 19 August 2013).
Cortina, J. M. (1993) 'What is coefficient alpha? An examination of theory and applications.', *Journal of Applied Psychology*, 78(1), pp. 98-104.
Crawford, E. G. C., Toft, Y. and Kift, R. L. (2013) 'New control room technologies: human factors analytical tools for railway safety', *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 227(5), pp. 529-538.
Creswell, J. W. (2014) *Research Design: Qualitative, Quantitative and Mixed Methods Approaches (4th ed.)*. London: SAGE Publications Ltd.
Cronbach, L. J. (1951) 'Coefficient Alpha and The Internal Structure of Tests', *Psychometrika*, 16(3), pp. 297-334.
Crotty, M. (1998) *The Foundations of Social Research: Meaning and Perspective in the Research Process*. London: Sage.
Crundall, D. E., Underwood, G. and Chapman, P. R. (1998) 'Chapter 18 - How Much Do Novice Drivers See? The Effects of Demand on Visual Search Strategies in Novice and Experienced Drivers', in *Eye Guidance in Reading and Scene Perception*. Amsterdam: Elsevier Science Ltd, pp. 395-417.
Cui, Y., Zhang, Y., Ma, Y. and Feng, Y. (2007) *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*. Sept. 30 2007-Oct. 3 2007.
Cullen (2000) *The Ladbroke Grove Rail Inquiry. Part 1 Report*. London: Commission, H. a. S.
Cynk, S., Tong, S., Kinnear, T., Weare, A., Delmonte, E., Hutchins, R. and Parkes, A. (2015) *Qualitative Study of Platform-Train Interface Incidents*. Wokingham.
Davies, D. R., Matthews, G., Stammers, R. B. and Westerman, S. J. (2013) *Human Performance: Cognition, Stress and Individual Differences*. Philadelphia: Taylor & Francis.
Davies, K., Hinton, J., Thorley, P. and Buse, B. (2007) *Feasibility of Head-Up Displays in Driving Cabs*. London: RSSB.

De Ambroggi, M. and Trucco, P. (2011) 'Modelling and assessment of dependent performance shaping factors through Analytic Network Process', *Reliability Engineering & System Safety*, 96(7), pp. 849-860.
De Araújo Fernandes Jr, S., Stetner Antonietti, L., Saba, A., Paulino De Faria, A., Maculano Esteves, A., Tufik, S. and Túlio De Mello, M. (2013) 'The impact of shift work on brazilian train drivers with different chronotypes: A comparative analysis through objective and subjective criteria', *Medical Principles and Practice*, 22(4), pp. 390-396.
De Ceunynck, T., Arien, C., Brijs, K., Brijs, T., Van Vlierden, K., Kuppens, J., Van Der Linden, M. and Wets, G. (2015) 'Proactive Evaluation of Traffic Signs Using a Traffic Sign Simulator', *EJTIR*, 15(2), pp. 184-204.
De Egea, B. G., Holgado, P. C. and Suárez, C. G. (2013) 'Humanscan®:A software solution towards the management of human reliability in the rail industry', *4th International Conference on Rail Human Factors* London. CRC Press, pp. 718-724.
Dekker, S. W. A. (2001) 'The re-invention of human error', *Human Factors and Aerospace Safety*, 1(3), pp. 247-265.
Denton, G. G. (1976) 'The Influence of Adaptation on Subjective Velocity for an Observer in Simulated Rectilinear Motion', *Ergonomics*, 19(4), pp. 409-430.
Department for Transport (2016) *Light Rail and Tram Statistics: England 2015/16*. London: DfT.
Desu, M. M. (2012) *Sample Size Methodology*. Burlington: Elsevier Science.
Deubel, H. (2008) 'The Time Course of Presaccadic Attention Shifts', *Psychological Research*, 72(6), pp. 630-640.
Dorrian, J., Hussey, F. and Dawson, D. (2007) 'Train driving efficiency and safety: Examining the cost of fatigue', *Journal of Sleep Research*, 16(1), pp. 1-11.
Dray, P., Sutton, L. and Menter, P. (1999) 'Signals Passed at Danger (SPADs): A new insight ', *6th Annual Railway Safety Congress*.
Duh, H. B.-L., Parker, D. E. and Furness, T. A. (2004) 'An Independent Visual Background Reduced Simulator Sickness in a Driving Simulator', *Presence: Teleoperators and Virtual Environments*, 13(5), pp. 578-588.
Dutta, A., Fisher, D. L. and Noyce, D. A. (2004) 'Use of a driving simulator to evaluate and optimize factors affecting understandability of variable message signs', *Transportation Research Part F: Traffic Psychology and Behaviour*, 7(4–5), pp. 209-227.
Dzeng, R.-J., Lin, C.-T. and Fang, Y.-C. (2016) 'Using eye-tracker to compare search patterns between experienced and novice workers for site hazard identification', *Safety Science*, 82, pp. 56-67.
Edquist, J., Horberry, T., Hosking, S. and Johnston, I. (2011) 'Effects of advertising billboards during simulated driving', *Applied Ergonomics*, 42(4), pp. 619-626.
Ehmke, C. and Wilson, C. (2007) *21st British HCI Group Annual Conference on HCI*.
Elliott, A. C. (2010) 'Human factors for railway signalling and control systems', *IET Professional Development Course on Railway Signalling and Control Systems*. Birmingham. RSCS, pp. 237-249.
Elliott, A. C. (2012) 'Human factors for railway signalling and control systems', *IET Professional Development Course on Railway Signalling and Control Systems*. London. RSCS, pp. 236-248.
Eugene, W., Dunbar, J., Nolan, A., Gilbert, J. E. and Hendrix, R. L. (2015) 'Designing for the Naturalistic Driving Experience', in Marcus, A. (ed.) *Design, User Experience, and Usability: Interactive Experience Design: 4th International Conference, DUXU 2015, Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, Proceedings, Part III*. Cham: Springer International Publishing,  pp. 439-449.

European Transport Safety Council (2003) *Transport Safety Performance in the EU. A Statistical Overview*. Brussels Council, E. T. S.
Evans, A. W. (2005) 'Railway risks, safety values and safety costs', *Proceedings of the Institution of Civil Engineers: Transport*, 158(1), pp. 3-9.
Feilzer, Y. M. (2010) 'Doing Mixed Methods Research Pragmatically: Implications for the Rediscovery of Pragmatism as a Research Paradigm', *Journal of Mixed Methods Research*, 4(1), pp. 6-16.
Fenner, D. (2002) 'Train protection', *IEE Review*, 48(5), p. 29.
Fereidunian, A., Zamani, M. A., Fatah, B., Lesani, H., Lucas, C., Kharazmi, P. and Torabi, H. (2010) *Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on*. 10-13 Oct. 2010.
Field, A. (2013) *Discovering Statistics Using IBM SPSS Statistics (4th ed.)*. London: SAGE Publications Ltd.
Finstad, K. (2010) 'Response Interpolation and Scale Sensitivity: Evidence Against 5-Point Scales', *Journal of Usability Studies*, 5(3), pp. 104-110.
Fitzpatrick, K., Chrysler, S., Park, E. S., Nelson, A., Robertson, J. and Iragavarapu, V. (2010) *Driver Workload at Higher Speeds*. College Station: Institute, T. T.
Folkard, S. (1997) 'Black times: Temporal determinants of transport safety', *Accident Analysis & Prevention*, 29(4), pp. 417-430.
Fowler, F. J. (2009) *Survey Research Methods (4th ed.)*. Washington: SAGE Publication, Inc.
Fujino, H., Horishita, T., Sonoda, T. and Yamaguchi, H. (2013) *4th International Conference on Rail Human Factors*. London.
Gardner, W., Mulvey, E. P. and Shaw, E. C. (1995) 'Regression Analyses of Counts and Rates: Poisson, Overdispersed Poisson and Negative Binomial Models', *Psychological Bulletin*, 118, pp. 392-404.
Gibson, H. (2016) *Industry Human Factors SPAD Review* London: RSSB.
Gibson, H., Mills, A. and Hesketh, S. (2009) *3rd International Conference on Rail Human Factors*. Lille.
Gibson, H., Mills, A., Smith, S. and Kirwan, B. K. (2013a) 'Railway action reliability assessment, a railway-specific approach to human error quantification', *4th International Conference on Rail Human Factors* London. CRC Press, pp. 671-676.
Gibson, H., Smith, S., Lowe, E., Mills, A., Morse, G. and Carpenter, S. (2013b) 'Incident factor classification system', *4th International Conference on Rail Human Factors* London. CRC Press, pp. 653-658.
Gillham, B. (2000) *Developing a Questionnaire*. London: Continuum.
Goldberg, J. H. and Kotval, X. P. (1999) 'Computer Interface Evaluation Using Eye Movements: Methods and Construct', *International Journal of Industrial Ergonomics*, 24(6), pp. 631-645.
Gourlay, C., Cole, C. and Rakotonirainy, A. (2013) 'Special Issue on work of the Cooperative Research Centre for Rail Innovation, Australia', *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 227(5), pp. 405-406.
Groeger, J. A., Bradshaw, M. F., Everett, J., Merat, N. and Field, D. (2001) *Pilot of Train Drivers Eye Movements*. Surrey, U. o.
Hale, A. R., Guldenmund, F. W., van Loenhout, P. L. C. H. and Oh, J. I. H. (2010) 'Evaluating safety management and culture interventions to improve safety: Effective intervention strategies', *Safety Science*, 48(8), pp. 1026-1035.
Halliday, M. W. (1995) *A human factors examination of wrong-side door release on Thames Trains*. Derby: Board, B. R.
Harris, R. L. and Christhilf, D. M. (1980).

Hauland, G. (2008) 'Measuring Individual and Team Situation Awareness During Planning Tasks in Training of En Route Air Traffic Control', *The International Journal of Aviation Psychology*, 18(3), pp. 290-304.
Hazarika, N. (2013) 'Correlation and Data Transformations', *Research Archives*. Available at: https://blog.majestic.com/research/correlation-data-transformations/.
Henderson, J. M. and Hollingworth, A. (1999) 'High-Level Scene Perception', *Annual Review of Psychology*, 50(1), pp. 243-271.
Henderson, J. M., Shinkareva, S. V., Wang, J., Luke, S. G. and Olerajczyk, J. (2013) 'Predicting Cognitive State from Eye Movements', *PLoS One*, 8(5).
Henderson, J. M., Weeks, P. A. and Hollingworth, A. (1999) 'The Effects of Semantic Consistency on Eye Movements During Complex Scene Viewing', *Journal of Experimental Psychology: Human Perception and Performance*, 25(1), pp. 210-228.
Herriotts, P. (2005) 'Identification of vehicle design requirements for older drivers', *Applied Ergonomics*, 36(3), pp. 255-262.
Hessels, R. S., Cornelissen, T. H. W., Kemner, C. and Hooge, I. T. C. (2015) 'Qualitative tests of remote eyetracker recovery and performance during head rotation', *Behavior Research Methods*, 47(3), pp. 848-859.
Hitchcock, D., Morris, C. and Taylor, A. (2013) 'A driver-centred driver's safety device', *4th International Conference on Rail Human Factors* London. CRC Press, pp. 109-117.
Ho, G., Scialfa, C. T., Caird, J. K. and Graw, T. (2001) 'Visual Search For Traffic Signs: The Effect of Clutter, Luminance, and Aging', *2001*, 43(2), pp. 194-207.
Hollnagel, E. (1998) *Cognitive reliability and error analysis method: CREAM*. London: Elsevier Science Ltd.
Holmqvist, K., Nystrom, M., Andersson, R., Dewhurst, R., Jarodzka, H. and Van Der Weijer, J. (2011) *Eye Tracking: a Comprehensive Guide to Methods and Measures*. New York: Oxford University Press.
Hooge, I. T. C., Vlaskamp, B. N. S. and Over, E. A. B. (2007) 'Saccadic Search: On the Duration of a Fixation', in Van Gompel, R., Fischer, M., Murray, W. S. and Hill, R. W. (eds.) *Eye Movement Research: Insights into Mind and Brain*. Oxford: Elsevier, pp. 581-595.
Horishita, T., Yamaura, K. and Kanayama, M. (2013) 'Study of effective praise in train driver's workplace', *4th International Conference on Rail Human Factors* London. CRC Press, pp. 614-620.
Hosking, S. G., Liu, C. C. and Bayly, M. (2010) 'The visual search patterns and hazard responses of experienced and inexperienced motorcycle riders', *Accident Analysis & Prevention*, 42(1), pp. 196-202.
Howard, D. F. (1976) 'TYNE AND WEAR METRO - A MODERN RAPID TRANSIT SYSTEM', *Inst Mech Eng (Lond) Proc*, 190(18), pp. 121-136.
Howard, J. V. (1998) 'The 2x2 Table: A Discussion from Bayessian Viewpoint', *Statistical Science*, 13(4), pp. 351-367.
Human Engineering Ltd (2005) *Analysis of the "May Peak" in SPAD Data*. London: RSSB.
Human Engineering Ltd (2006) *Investigating the Design of a Display to Repeat Signal Aspects* (T350). London: RSSB.
Huson (2007) 'Performance of Some Correlation Coefficients When Applied to Zero-Clustered Data', *Journal of Modern Applied Statistical Methods*, 6(2), pp. 530-536.
Inoue, D. and Yamamoto, S. (2006) 'Support for drivers via just-in-time predictive control and fault detection based on a nearest neighbor method during braking to stop trains', *Nihon Kikai Gakkai Ronbunshu, C Hen/Transactions of the Japan Society of Mechanical Engineers, Part C*, 72(9), pp. 2756-2761.

Inoue, T., Kitamura, Y., Sato, A. and Masuda, T. (2015) 'Coaching technique for safety training using self-assessment', *Quarterly Report of RTRI (Railway Technical Research Institute)*, 56(1), pp. 51-58.
Institution of Occupational Safety and Health (2016) *Managing Safely (Version 4.0)*. Wigston.
Jacob, R. J. K. and Karn, K. S. (2003) 'Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises', in Radach, R., Hyona, J. and Deubel, H. (eds.) *The Mind's Eye*. Oxford: Elsevier Science BV.
Jamieson, S. (2004) 'Likert scales: how to (ab)use them', *Medical Education*, 38(12), pp. 1217-1218.
Jamson, A. H. and Motula, S. (2004) *Driving Simulator Conference*. DSC.
Jamson, S. L. and Jamson, A. H. (2010) 'The validity of a low-cost simulator for the assessment of the effects of in-vehicle information systems', *Safety Science*, 48(10), pp. 1477-1483.
Jang, Y.-M., Mallipeddi, R., Lee, S., Kwak, H.-W. and Lee, M. (2014) 'Human intention recognition based on eyeball movement pattern and pupil size variation', *Neurocomputing*, 128, pp. 421-432.
Jeon, H. J., Kim, J. H., Kim Ma, B. N., Park, S. J., Fava, M., Mischoulon, D., Choi, H., Roh, S. and Lee, D. (2014) 'Sleep quality, posttraumatic stress, depression, and human errors in train drivers: A population-based nationwide study in South Korea', *Sleep*, 37(12), pp. 1969-1975.
Johanning, E. (2011) 'Vibration and shock exposure of maintenance-of-way vehicles in the railroad industry', *Applied Ergonomics*, 42(4), pp. 555-562.
Johnson, T. M., Hanson, C. E., Ross, J. C. and Zaouk, A. K. (2009) 'Development of passive and active noise control for next generation locomotive cabs', *38th International Congress and Exposition on Noise Control Engineering 2009*. Ottawa. INTER-NOISE, pp. 927-935.
Jolly, L., Kavanagh, L., O'Moore, L. and Tibbits, G. (2013) *Evaluation of Simulators - Final Report*. Brisbane.
Just, M. A. and Carpenter, P. A. (1976) 'Eye fixations and cognitive processes', *Cognitive Psychology*, 8(4), pp. 441-480.
Just, M. A. and Carpenter, P. A. (1980) 'A Theory of Reading: From Eye Fixations to Comprehension', *Psychological Review*, 87(4), pp. 329-354.
Kaiser, M. K. and Schroeder, J. A. (2003) 'Flights of Fancy: The Art and Science of Flight Simulation', in P.S., T. and Vidulich, M. A. (eds.) *Principles and Practice of Aviation Psychology*. Mahwah, New Jersey: Lawrense Erlbaum Associates,  pp. 458-465.
Kapitaniak, B., Walczak, M., Kosobudzki, M., Jozwiak, Z. and Bortkiewicz, A. (2015) 'Application of eye-tracking in drivers testing: A review of research', *International Journal of Occupational Medicine and Environmental Health* 28(6), pp. 941-954.
Kappé, B., van Erp, J. and Korteling, J. E. (1999) 'Effects of Head-Slaved and Peripheral Displays on Lane-Keeping Performance and Spatial Orientation', *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 41(3), pp. 453-466.
Karowich, P., Mallett, C., Woods, C. and Cowie, H. (2009) 'Creating a culture of safe driving behaviors' *SPE Reprint Series* [Article].
Kasarkis, P., Stehwien, J., Hickox, J. C., Aretz, A. and Wickens, C. D. (2001) *11th International Symposium on Aviation Psychology*
Kecklund, L., Lindgrenwalter, A. and Nordlöf, E. (2013) 'Investigating the effects of the ERTMS driver machine interface on train driver behaviour and railway safety', *4th International Conference on Rail Human Factors* London. CRC Press, pp. 201-205.

Keun Sang, P. and Ohkubo, T. (1994) 'A case study on the relationship between neuro-sensory work and work load', *Computers and Industrial Engineering*, 27(1-4), pp. 393-396.

Kim, J. W. and Jung, W. (2003) 'A taxonomy of performance influencing factors for human reliability analysis of emergency tasks', *Journal of Loss Prevention in the Process Industries*, 16(6), pp. 479-495.

Knapper, A., Christoph, M., Hagenzieker, M. and Brookhuis, K. (2015) 'Comparing a driving simulator to the real road regarding distracted driving speed', *European Journal of Transport and Infrastructure Research*, 15(2), pp. 205-225.

Kontogiannis, T. and Malakis, S. (2009) 'A proactive approach to human error detection and identification in aviation and air traffic control', *Safety Science*, 47(5), pp. 693-706.

Kostoulas, A. (2013) *Four Things You Probably Didn't Know About Likert Scales*. Available at: http://achilleaskostoulas.com/2013/09/09/four-things-you-probably-didnt-know-about-likert-scales/ (Accessed: 16 January 2015).

Kowalski, C. J. (1972) 'On the Effects of Non-Normality on the Distribution of the Sample Product-Moment Correlation Coefficient', *Journal of Royal Statistical Society. Series C (Applied Statistics)*, 21(1), pp. 1-12.

Kyriakidis, M. (2013) 'A framework to identify and quantify the most significant performance shaping factors in railway operations ' [PowerPoint Slides]. RRUKA Next Generation Rail Event. Available at: http://www.sparkrail.org/Lists/Records/DispForm.aspx?ID=8700.

Kyriakidis, M., Hirsch, R. and Majumdar, A. (2012a) 'Metro railway safety: An analysis of accident precursors', *Safety Science*, 50(7), pp. 1535-1548.

Kyriakidis, M., Majumdar, A., Grote, G. and Ochieng, W. Y. (2012b) 'Development and assessment of taxonomy for performance-shaping factors for railway operations'. pp. 145-153.

Kyriakidis, M., Majumdar, A. and Ochieng, W. Y. (2012c) 'A Human Performance Operational Railway Index to Estimate Operator's Error Probability', in Stanton, N. A. (ed.) *Advances in Human Aspects of Road and Rail Transportation*. CRC Press, pp. 832-841.

Kyriakidis, M., Majumdar, A. and Ochieng, W. Y. (2015) 'Data Based Framework to Identify the Most Significant Performance Shaping Factors in Railway Operations', *Safety Science*, 78, pp. 60-76.

Large, D. R., Golightly, D. and Taylor, E. (2015) *5th International Rail Human Factors Conference*. London.

Large, D. R., Golightly, D. and Taylor, E. L. (2014) 'The effect of driver advisory systems on train driver workload and performance', in *Contemporary Ergonomics and Human Factors 2014*. Taylor & Francis, pp. 335-342.

LaViola, J. J. J. (2000) 'A discussion of cybersickness in virtual environments', *SIGCHI Bull.*, 32(1), pp. 47-56.

Lawton, R. and Ward, N. J. (2005) 'A systems analysis of the Ladbroke Grove rail crash', *Accident Analysis and Prevention*, 37(2), pp. 235-244.

Lemieux, C., Stinchcombe, A., Gagnon, S. and Bédard, M. (2014) 'Comparison of simulated driving performance across platforms: From ‘low-cost desktop’ to ‘mid-level’ driving simulators', *Advances in Transportation Studies*, 34, pp. 33-42.

Lewis, J. and Sauro, J. (2009) 'The Factor Structure of the System Usability Scale', in Kurosu, M. (ed.) *Human Centered Design*. Springer Berlin Heidelberg, pp. 94-103.

Li, G. (2003) *Railway signal sighting: Scientific evidence for sighting principles*. Bristol: Ltd, H. E.

Li, G. (2004) *Analysis of the 'May Peak' in SPAD data: Human Factors support to SPAD management in the Southern and Scotland Zones - Summary Report (T255)*. London: RSSB.

Li, G., Hamilton, W. I., Morrisroe, G. and Clarke, T. (2006) 'Driver detection and recognition of lineside signals and signs at different approach speeds', *Cognition, Technology and Work*, 8(1), pp. 30-40.

Lindner, T. (2013) 'The influence of selecting a specific sample of hazardous events on the evaluation of the train driver's reliability', *4th International Conference on Rail Human Factors* London. CRC Press, pp. 700-706.

Macdonald, I., Smith, L., Lowe, S. L. and Folkard, S. (1997) 'Effects on accidents of time into shift and of short breaks between shifts', *International Journal of Occupational and Environmental Health*, 3(3 SUPPL.), pp. S40-S45.

Mackieh, A. and Cilingir, C. (1998) 'Effects of performance shaping factors on human error', *International Journal of Industrial Ergonomics*, 22(4-5), pp. 285-292.

Madigan, R., Golightly, D. and Madders, R. (2016) 'Application of Human Factors Analysis and Classification System (HFACS) to UK rail safety of the line incidents', *Accident Analysis & Prevention*, 97, pp. 122-131.

Maguire, D. J. (2009) 'Active noise reduction and vibration control in locomotive cabs - Issues, impacts, and motivations for rail operations worldwide', *16th International Congress on Sound and Vibration 2009*. Krakow. Curran Associates, Inc, pp. 2919-2926.

Mangan, J., Lalwani, C. and Gardner, B. (2004) 'Combining Quantitative and Qualitative Methodologies in Logistics Research', *International Journal of Physical Distribution & Logistics Management*, 34(7), pp. 565-578.

Marquart, G., Cabrall, C. and de Winter, J. (2015) 'Review of Eye-Related Measures of Drivers' Mental Workload ', *Procedia Manufacturing* 3, pp. 2854-2861.

Marques de Sa, J. P. (2007) *Applied Statistics Using SPSS, STATISTICA, MATLAB and R*. Berlin: Springer-Verlag.

Martens, M. H. (2000) 'Assessing Road Sign Perception: A Methodological Review', *Transportation Human Factors*, 2(4), pp. 347-357.

Masaaki, I., Yamakawa, T., Amakusa, M., Kuratani, M., Takeda, Y. and Kusukami, K. (2015) 'Development of a human factors education programme for train drivers combined with simulator training ', *Fifth International Rail Human Factors Conference*. London.

Megaw, E. D. and Richardson, J. (1979) 'Eye Movements and Industrial Inspection', *Applied Ergonomics*, 10(3), pp. 145-154.

Milleville-Pennel, I., Pothier, J., Hoc, J.-M. and Mathé, J.-F. (2010) 'Consequences of cognitive impairments following traumatic brain injury: Pilot study on visual exploration while driving', *Brain Injury*, 24(4), pp. 678-691.

Mirowska, M. and Mroz, E. (2000) 'Effect of low frequency noise at low levels on human health in light of questionnaire investigation', *Internoise*. Nice. pp. 2809-2812.

Miura, T. (1990) 'Active Function of Eye Movement and Useful Field of View in a Realistic Setting', in Groner, R., D'YDewalle, G. and Parham, R. (eds.) *From Eye to Mind: Information Acquisition in Perception, Search and Reading* Amsterdam: Elsevier Science Publishers, pp. 119-127.

Monk, A., Johnson, R., Murphy, P., Mills, A., Dacre, M. and Cassidy, S. (2015) 'Are drivers complying with speed restrictions? ', *Fifth International Rail Human Factors Conference*. London.

Montano, G. (2011) *Dynamic Reconfiguration of Safety-Critical Systems: Automation and Human Involvement*. University of York.

Moore, D. S., McCabe, G. P. and Craig, B. A. (2009) *Introduction to the Practice of Statistics (6th ed.)*. New York: W.H. Freeman.

Moors, A., O'Connell, P., Mills, A. and Stacy, B. (2015) *Fifth International Rail Human Factors Conference*. London.
Morrel-Samuels, P. (2002) 'Getting the Truth into Workplace Surveys', *Harvard Business Review*.
Multer, J., Safar, H. and Roth, E. (2015) *Fifth International Rail Human Factors Conference*. London.
Myrtek, M., Deutschmann-Janicke, E., Strohmaier, H., Zimmermann, W., Lawerenz, S., BrÜGner, G. and MÜLler, W. (1994) 'Physical, mental, emotional, and subjective workload components in train drivers', *Ergonomics*, 37(7), pp. 1195-1203.
Naghiyev, A., Sharples, S., Carey, M., Coplestone, A. and Ryan, B. (2014) 'ERTMS train driving-in cab vs. outside: An explorative eye-tracking field study', *Contemporary Ergonomics and Human Factors 2014*. pp. 343-350.
Naghiyev, A., Sharples, S., Ryan, B., Coplestone, A. and Carey, M. (2015) 'Evaluating existing train driving models to generate guidance for future ERTMS train driving models ', *Fifth International Rail Human Factors Conference*. London.
NaturalPoint Inc (2016) *TrackIR5 - In Depth*. Available at: https://www.naturalpoint.com/trackir/trackir5/ (Accessed: 04/09/16).
Naumann, A., Grippenkoven, J. and Lemmer, K. (2016) 'Future Information and Assistance Systems for Train Drivers and Evalulation of Their Usability', *World Congress on Railway Research 2016*. Milan.(Accessed: 22/08/2016).
Naweed, A. (2013) 'Simulator integration in the rail industry: the Robocop problem', *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 227(5), pp. 407-418.
Naweed, A. and Balakrishnan, G. (2013) 'That train has already left the station! Improving the fidelity of a railway safety research simulator at post-deployment', *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 227(5), pp. 419-426.
Naweed, A. and Balakrishnan, G. (2014) 'Understanding the visual skills and strategies of train drivers in the urban rail environment.', *Work*, 47(3).
Naweed, A. and Rainbird, S. (2014) *SPAD-risk management & mitigation*. Brisbane: Innovation, C. f. R.
ND BOMEL Ltd (2009) *Criteria for the Application of Selective Door Operation. Final Report*. London: RSSB.
Neisser, U. (1976) *Cognition and reality: principles and implications of cognitive psychology*. San Francisco: W.H. Freeman.
Norman, D. A. (1981) 'Categorization of action slips', *Psychological Review*, 88(1), pp. 1-15.
Norman, D. A. (1988) *The psychology of everyday things*. New York: Basic Books.
Nunnally, J. C. and Bernstein, I. H. (1994) *Psychometric theory (3rd ed.)*. New York: McGraw-Hill.
O'Connell, P., Lawton, F. and Mills, A. (2015) 'DOO Tell: Developing risk triggered commentary driving and stabilised approach procedures for driver only operations in New Zealand. ', *Fifth International Rail Human Factors Conference*. London.
Office of the Chief Investigator (2009) *Rail Safety Investigation Report No 2009/05 - Platform Overruns. Siemens Nexas EMU* (2009/05). Melbourne.
Olsen, A. (2012) *The Tobii I-VT Fixation Filter*. Technology, T.
Ottati, W. L., Hickox, J. C. and Richter, J. (1999) 'Eye Scan Patterns of Experienced and Novice Pilots during Visual Flight Rules (VFR) Navigation', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 43(1), pp. 66-70.
Parasuraman, R., Sheridan, T. B. and Wickens, C. D. (2000) 'A model for types and levels of human interaction with automation', *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 30(3), pp. 286-297.

Parkes, A. M. and Hooijmeijer, V. (2001) *1st Human-Centered Transportation Simulation Conference*. Iowa.
Patton, M. Q. (2002) *Qualitative Research and Evaluation Methods (3rd Ed.)*. Thousand Oaks: Sage.
Philips, D. C. and Burbules, N. C. (2000) *Postpositivism and Educational Research*. Lanham, NY: Rowman & Littlefield.
Pitsopoulos, J., Pitsopoulos, C. and Regan, M. A. (2010) *Development of a Train Driver Education Programme on Mobile Phone Risk (T904)*. London.
Plant, K. L. and Stanton, N. A. (2012) 'Why did the pilots shut down the wrong engine? Explaining errors in context using Schema Theory and the Perceptual Cycle Model', *Safety Science*, 50(2), pp. 300-315.
Poole, A., Ball, L. J. and Phillips, P. (2004) *HCI Conference on People and Computers XVIII*. London. Springer-Verlar.
Powell, J. P., González-Gil, A. and Palacin, R. (2014) 'Experimental assessment of the energy consumption of urban rail vehicles during stabling hours: Influence of ambient temperature', *Applied Thermal Engineering*, 66(1–2), pp. 541-547.
Powell, J. P. and Palacín, R. (2015) 'Passenger Stability Within Moving Railway Vehicles: Limits on Maximum Longitudinal Acceleration', *Urban Rail Transit*, 1(2), pp. 95-103.
Pradhan, A. K., Hammel, K. R., DeRamus, R., Pollatsek, A., Noyce, D. A. and Fisher, D. L. (2005) 'Using Eye Movements To Evaluate Effects of Driver Age on Risk Perception in a Driving Simulator', *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 47(4), pp. 840-852.
Profillidis, V. A. (2000) *Railway engineering*. 2nd ed.. edn. Aldershot: Aldershot : Avebury Technical.
Prottengeier, J., Petzoldt, M., Jess, N., Moritz, A., Gall, C., Schmidt, J. and Bauer, G. (2016) 'The effect of a standardised source of divided attention in airway management: A randomised, crossover, interventional manikin study.', *European Journal of Anaesthesiology*, 33(3), pp. 195-203.
RAIB (2012) *Fatal accident at James Street station, Liverpool 22 October 2011*. Derby: RAIB.
RAIB (2013) *Accident at Charing Cross Station 24 November 2012*. Derby.
RAIB (2016a) *Passenger Trapped and Dragged by a Train at Hayes & Harlington Station*. Derby.
RAIB (2016b) *Passenger Trapped in Train Doors and Dragged at Clapham South Station*. Derby.
Railway Group Standards (2000) *GI/RT7006 - Prevention and mitigation of overruns*. London: Safety & Standards Directorate
Railway Group Standards (2003) *GE/GN8537: Guidance on Signal Positioning and Visibility*. London: RSSB.
Railway Group Standards (2009) *GI/RT7033 Lineside Operational Safety Signs*. London: RSSB.
Rajkumar, D., Sanjar, A. S. and Dhamodharan, R. (2012) *Platform–train interface for rail passengers – Technology review* Brisbane.
Rasmussen, J. (1983) 'Skills, rules, and knowledge; Signals, signs, symbols, and other destinctions in human performance models', *IEEE Transactions on Systems, Man and Cybernetics*, SMC-13(3), pp. 257-266.
Rayner, K. (1998) 'Eye Movements in Reading and Information Processing: 20 Years of Research', *Psychological Bulletin*, 124(4), pp. 372-422.
Reason, J. (1990) *Human Error*. Cambridge: Cambridge University Press.
Reason, J. (1997) *Managing the Risks of Organizational Accidents*. Aldershot: Ashgate Publishing.

Reason, J., Manstead, A., Stephen, S., Baxter, J. and Campbell, K. (1990) 'Errors and violations on the roads: A real distinction?', *Ergonomics*, 33(10-11), pp. 1315-1332.
Reed, M. P. and Green, P. A. (2010) 'Comparison of driving performance on-road and in a low-cost simulator using a concurrent telephone dialling task', *Ergonomics*, 42(8), pp. 1015-1037.
Reingold, E. M. and Charness, N. (2005) 'Perception In Chess: Evidence From Eye Movements', in Underwood, G. M. (ed.) *Cognitive Processes in Eye Guidance*. Oxford: Oxford University Press, pp. 325-354.
Reingold, E. M., Charness, N., Pomplun, M. and Stampe, D. M. (2001) 'Visual Span in Expert Chess Players: Evidence From Eye Movements', *Psychological Science*, 12(1), pp. 48-55.
Rentzsch, M., Seliger, D., Miglianico, D. and Georget, D. (2009) *3rd International Conference on Rail Human Factors*. Lille.
RIAC (2007) *SPADs and Human Factors*. London: ORR.
Roels, R. and Mills, A. (2010) *Causes, risks and mitigation of station stopping incidents: Project summary report* Londong: RSSB.
Roth, E. M., Multer, J. and Raslear, T. (2006) 'Shared Situation Awareness as a Contributor to High Reliability Performance in Railroad Operations', *Organization Studies*, 27(7), pp. 967-987.
Rowe, I. (2013) 'Developing a tram driver route learning training simulator for manchester's metrolink trams', *4th International Conference on Rail Human Factors* London. CRC Press, pp. 262-269.
RSSB (2004a) *Impact of European Rail Traffic Management System on driver workload*. London.
RSSB (2004b) *T070 - Common Factors in SPADs*. London: RSSB.
RSSB (2005) *T152 - Train Driver Visual Strategies*. London: RSSB.
RSSB (2006) *Development of SPAD risk management tools - summary report*. London: RSSB.
RSSB (2007a) *Human Factors of CCTV monitoring*. London.
RSSB (2007b) *Understanding multi-SPAD signals and the effects of line identifiers on driver behaviour: an investigation into multi-SPAD signal D301.* (T632). London: RSSB.
RSSB (2008) *Understanding Human Factors: a guide of the railway industry*. London: RSSB.
RSSB (2010) *Causes, risk and mitigation of station stopping incidents (T595)* London: RSSB.
RSSB (2011a) *Passenger risk at the platform-train interface*. London: RSSB.
RSSB (2011b) *Precursor Indicator Model (PIM) for Train Accident Risk*. London: (Accessed: 04/09/2016).
RSSB (2012) *ERTMS/ETCS driver/machine interface options for future train cab designs. New build system design assessment and options report*. London.
RSSB (2013) *SPAD and TPWS activity report. Quarter 2 - 2013/14*. London: RSSB.
RSSB (2014a) *Application of Image Recognition Software at the Platform-Train Interface*. London.
RSSB (2014b) *Evaluating technological solutions to support driver only operation train dispatch (T1035)* London.
RSSB (2016) 'Annual Safety Performance Report 2015-2016'. London: RSSB.
RSSB, Halcrow Group Ltd and Human Engineering Ltd (2006) *T426 - Minimisation of Accidents at the Platform Train Interface*. London: RSSB.
Rushby, N. and Seabrook, J. (2007) *Recent Advances in Simulation Training and Assessment for the Rail Industry: Results and Case Studies (T711)*. London.

Russell, A. J., Bailey, S. and Moore, A. C. (2013) 'The introduction of non-technical skills in to a train driver competence management system', *4th International Conference on Rail Human Factors* London. CRC Press, pp. 529-538.
Ružić, D. and Časnji, F. (2011) 'Agricultural tractor cab characteristics relevant for microclimatic conditions', *Journal of Applied Engineering Science*, 9(2), pp. 323-330.
Ryan, B., Hutchings, J. and Lowe, E. (2010) 'An analysis of the content of questions and responses in incident investigations: Self reports in the investigation of signals passed at danger (SPADs)', *Safety Science*, 48(3), pp. 372-381.
Ryan, B., Wilson, J. R., Sharples, S. and Clarke, T. (2009a) 'Attitudes and opinions of railway signallers and related staff, using the Rail Ergonomics Questionnaire (REQUEST)', *Applied Ergonomics*, 40(2), pp. 230-238.
Ryan, B., Wilson, J. R., Sharples, S., Morrisroe, G. and Clarke, T. (2009b) 'Developing a Rail Ergonomics Questionnaire (REQUEST)', *Applied Ergonomics*, 40(2), pp. 216-229.
Salmon, P. M., Goode, N., Stevens, E., Walker, G. and Stanton, N. A. 9174 (2015) 'The elephant in the room: Normal performance and accident analysis' *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* [Conference Paper]. pp. 275-285.
Salmon, P. M., Young, K. L. and Regan, M. A. (2011) 'Distraction ‘on the buses’: A novel framework of ergonomics methods for identifying sources and effects of bus driver distraction', *Applied Ergonomics*, 42(4), pp. 602-610.
Santiago-Chaparro, K. R., DeAmico, M., Bill, A., Chitturi, M. and Noyce, D. (2011) 'Realistic-Scenario Creation Process for Virtual Road Safety Audits', *Advances in Transportation Studies*, pp. 19-28.
Sato, A. and Bowler, N. (2015) *5th International Rail Human Factors Conference*. London.
Sato, H., Marumo, Y., Tsunashima, H. and Kojima, T. (2010) 'Braking assistance system for train drivers by indicating predicted stopping position', *SICE Annual Conference 2010*. Taipei. pp. 1353-1357.
Savelsbergh, G. J. P., Williams, A. M., Van Der Kamp, J. and Ward, P. (2002) 'Visual Search, Anticipation and Expertise In Soccer Goalkeepers', *Journal of Soprts Sciences*, 20(3), pp. 279-287.
Schulz, C. M., Schneider, E., Fritz, L., Vockeroth, J., Hapfelmeier, A., Wasmaier, M., Kochs, E. F. and Schneider, G. (2011) 'Eye tracking for assessment of workload: a pilot study in an anaesthesia simulator environment', *BJA: British Journal of Anaesthesia*, 106(1), pp. 44-50.
Shappel, S. and Wiegmann, D. (2000) *The Human Factors Analysis and Classification System - HFACS*. Washington: Transportation, U. S. D. o.
Shechtman, O. (2010) 'Validation of driving simulators', *Advances in Transportation Studies*, (SPEC. ISSUE), pp. 53-62.
Shimojo, S., Simion, C., Shimojo, E. and Scheier, C. (2003) 'Gaze Bias Both Reflects and Influences Preference', *Nature Neuroscience*, 6(12), pp. 1317-1322.
Shinar, D., McDowell, E. D. and Rockwell, T. H. (1977) 'Eye Movements in Curve Negotiation', *Human Factors*, 19(1), pp. 63-71.
Simmons, D. (2015) 'Virtual World Simulation in Systems Design', *Fifth International Rail Human Factors Conference*. London.
Sodhi, M., Reimer, B. and Llamazares, I. (2002) 'Glance analysis of driver eye movements to evaluate distraction', *Behavior Research Methods, Instruments, & Computers*, 34(4), pp. 529-538.
Stanton, N. A. and Salmon, P. M. (2009) 'Human error taxonomies applied to driving: A generic driver error taxonomy and its implications for intelligent transport systems', *Safety Science*, 47(2), pp. 227-237.

Stanton, N. A. and Walker, G. H. (2011) 'Exploring the psychological factors involved in the Ladbroke Grove rail accident', *Accident Analysis & Prevention*, 43(3), pp. 1117-1127.
Stevenson, M. G., Coleman, N., Long, A. F. and Williamson, A. M. (2000) 'Assessment, re-design and evaluation of changes to the driver's cab in a suburban electric train', *Applied Ergonomics*, 31(5), pp. 499-506.
Strauss, E., Sherman, E. M. S. and Spreen, O. (2006) *A Compendium of Neuropsychological Tests: Administration, Norms and Commentary (3rd ed.)*. New York: Oxford University Press.
Sturm, W., Willmes, K., Orgass, B. and Hartje, W. (1997) 'Do Specific Attention Deficits Need Specific Training?', *Neuropsychological rehabilitation*, 7(2), pp. 81-103.
Sümer, S. K., Say, S. M., Ege, F. and Sabanci, A. (2006) 'Noise exposed of the operators of combine harvesters with and without a cab', *Applied Ergonomics*, 37(6), pp. 749-756.
Sumpor, D., Tos, Z. and Musabašić, N. (2013) 'Static anthropometry measures of tram drivers in bosnia & herzegovina important fortram control panel design', *4th International Conference on Rail Human Factors*. London. CRC Press, pp. 118-125.
Swanson, J. D. (2004).
Szulewski, A. and Howes, D. (2014) 'Combining First-Person Video and Gaze-Tracking in Medical Simulation: A Technical Feasibility Study', *The Scientific World Journal*, 2014, p. 4.
Tatler, B. W. and Vincent, B. T. (2008) 'Systematic Tendencies in Scene Viewing', *Journal of Eye Movement Research*, 2(5), pp. 1-18.
Teddlie, C. and Tashakkori, A. (2009) *Foundations of mixed methods research : integrating quantitative and qualitative techniques in the social and behavioral sciences*. London: SAGE.
The European Rail Research Advisory Council (2009) *Metro, light rail and tram systems in Europe*. ERRAC.
Thiffault, P. and Bergeron, J. (2003) 'Monotony of road environment and driver fatigue: a simulator study', *Accident Analysis & Prevention*, 35(3), pp. 381-391.
Thomas, A. and Gassel, C. (2009) *3rd International Conference on Rail Human Factors*. Lille.
Törnros, J. (1998) 'Driving behaviour in a real and a simulated road tunnel—a validation study', *Accident Analysis & Prevention*, 30(4), pp. 497-503.
Traub, P. and Fraser, G. (2013) 'Safety assessment of CCTV for platform interface tasks - Sydney suburban trains', *4th International Conference on Rail Human Factors* London. CRC Press, pp. 300-308.
Tripathi, K. and Borrion, H. (2016) 'Safe, secure or punctual? A simulator study of train driver response to reports of explosives on a metro train', *Security Journal*, 29(1), pp. 87-105.
Tschirner, S., Andersson, A. W. and Sandblad, B. (2013) 'Designing train driver advisory systems for situation awareness', *4th International Conference on Rail Human Factors* London.
Tu, W. (2006) 'Zero-Inflated Data', in *Encyclopedia of Envirometrics*. John Wiley & Sons.
Underwood, G. (2007) 'Visual attention and the transition from novice to advanced driver', *Ergonomics*, 50(8), pp. 1235-1249.
Underwood, G., Crundall, D. and Chapman, P. (2002) 'Selective searching while driving: the role of experience in hazard detection and general surveillance', *Ergonomics*, 45(1), pp. 1-12.
Underwood, G., Ngai, A. and Underwood, J. (2013) 'Driving experience and situation awareness in hazard detection', *Safety Science*, 56, pp. 29-35.

Unema, P. J. A. and Rotting, M. (1990) 'Differences in Eye Movements and Mental Workload Between Experienced and Inexperienced Motor-Vehicle Drivers', in Brogan, D. (ed.) *Visual Search*. London: Taylor & Francis, pp. 193-202.
Van Der Weide, R., Frieling, H. F. L., Malle, F. and Miglianico, D. (2013) 'Amsterdam metro cab: Ergonomics in the design, verification and validation process', *4th International Conference on Rail Human Factors* London. CRC Press, pp. 270-279.
Van Luipen, J. and Meijer, S. (2009) *3rd International Conference on Rail Human Factors*. Lille.
Van Orden, K. F., Limbert, W., Makeig, S. and Jung, T. P. (2001) 'Eye Activity Correlates of Workload During a Visuospatial Memory Task', *Human Factors*, 43(1), pp. 111-121.
Vatshelle, Å. and Moen, B. E. (1997) 'Serious on-the-track accidents experienced by train drivers: Psychological reactions and long-term health effects', *Journal of Psychosomatic Research*, 42(1), pp. 43-52.
Verstappen, V. (2015) 'Driving performance of Dutch train drivers: Assessing the impact of the presence of a second person in the cabin ', *Fifth International Rail Human Factors Conference*. London.
Vitale, D. C., Armenakis, A. A. and Feild, H. S. (2008) 'Integrating qualitative and quantitative methods for organizational diagnosis: Possible priming effects?', *Journal of Mixed Methods Research*, 2(1), pp. 87-105.
Vives, J. and Losilla, J.-M. (2006) 'Count Data in Psychological Applied Research', *Psychological Reports*, 98, pp. 821-835.
Vogt, J., Leonhardt, J., Köper, B. and Pennig, S. (2010) 'Human factors in safety and business management', *Ergonomics*, 53(2), pp. 149-163.
Vuchic, V. R. (2007) *Urban transit systems and technology*. Hoboken, N.J. : Chichester: Hoboken, N.J. : Wiley
Chichester : John Wiley.
Ward, C., Goodall, R., Dixon, R. and Charles, G. (2012) 'Detection of low adhesion in the railway vehicle wheel/rail interface: Assessment of multi-bodied simulation data', *2012 UKACC International Conference on Control*. Cardiff. pp. 725-730.
Williamson, A., Lombardi, D. A., Folkard, S., Stutts, J., Courtney, T. K. and Connor, J. L. (2011) 'The link between fatigue and safety', *Accident Analysis & Prevention*, 43(2), pp. 498-515.
Wilson, J. R. and Norris, B. J. (2006) 'Human factors in support of a successful railway: A review', *Cognition, Technology and Work*, 8(1), pp. 4-14.
Woods, D., Dekker, S., Cook, R., Johannesen, L. and Sarter, N. (2010) *Behind Human Error*. Aldershot: Ashgate Publishing.
Yan, X., Abdel-Aty, M., Radwan, E., Wang, X. and Chilakapati, P. (2008) 'Validating a driving simulator using surrogate safety measures', *Accident Analysis & Prevention*, 40(1), pp. 274-288.
Yang, H. K., Lee, J. W., Lee, Y. J., Lee, J. H., Lim, M. G., Baek, J. H. and Song, Y. S. (2012) 'A Study Concerning Analysis of Arousal State of locomotive Engineering during Operating Train', *Transactions of the Korean Institute of Electrical Engineers*, 61(6), pp. 891-898.
Yarbus, A. L. (1967) *Eye Movements and Vision*. New York: Plenum Press.
Yates, D., Moore, D. S. and McCabe, G. P. (1999) *The Practice of Statistics*. New York: W.H.Freeman.
Yates, T. K. and Sharples, S. (2009) *3rd International Conference on Rail Human Factors*. Lille.
Young, K. L., Lenné, M. G., Beanland, V., Salmon, P. M. and Stanton, N. A. (2015) 'Where do novice and experienced drivers direct their attention on approach to urban rail level crossings?', *Accident Analysis & Prevention*, 77, pp. 1-11.

Young, M. S. and Grenier, D. P. J. (2009) 'Future Train Cab Interface Design: Development of a Model to Support Driver Situation Awareness', *3rd International Conference on Rail Human Factors*. Lille.

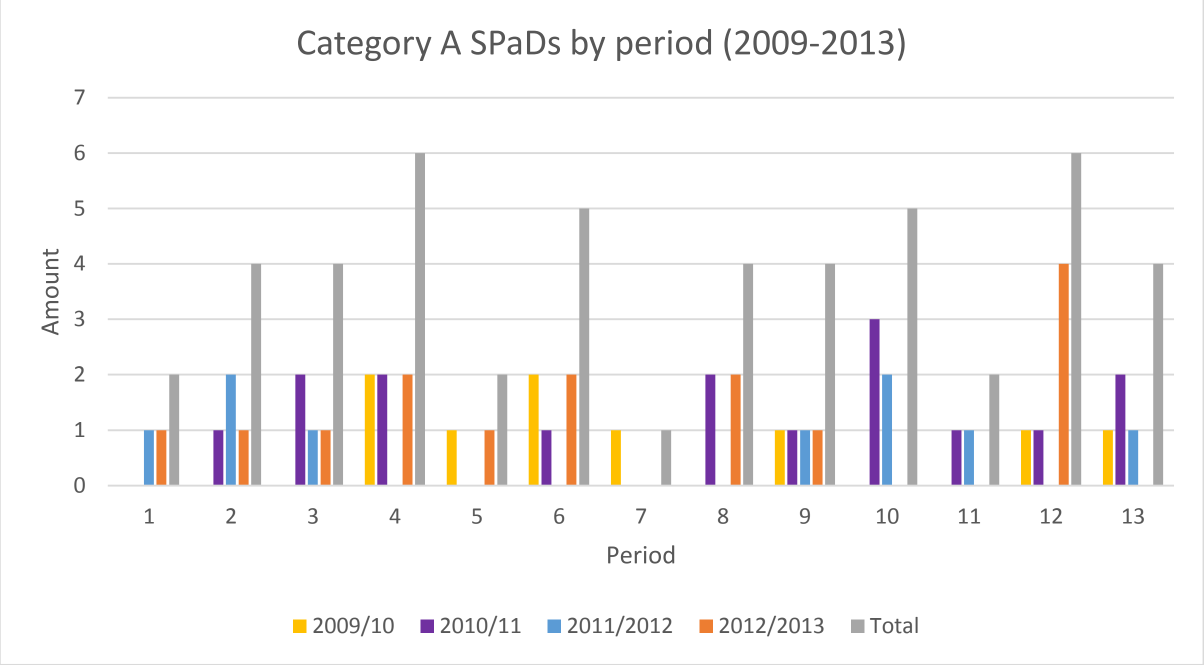
Yum, B. S., Roh, J. H., Ryu, J. C., Won, J. U., Kim, C. N., Lee, J. E. and Kim, K. Y. (2006) 'Symptoms of PTSD according to individual and work environment characteristics of Korean railroad drivers with experience of person-under-train accidents', *Journal of Psychosomatic Research*, 61(5), pp. 691-697.

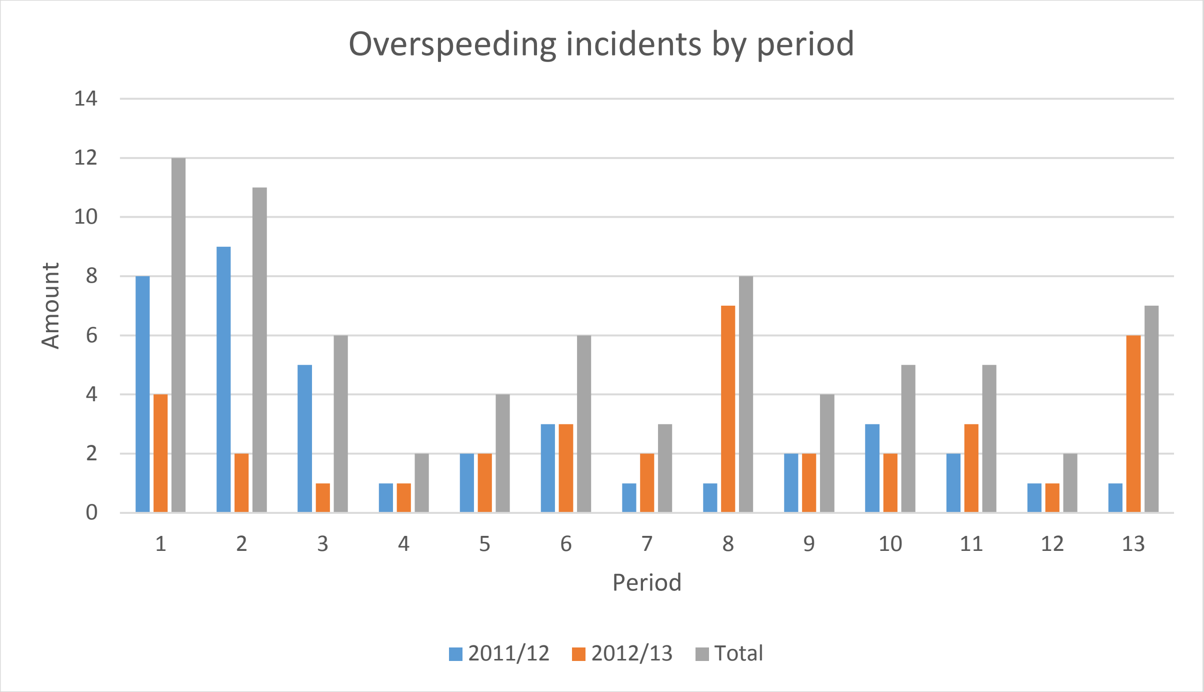
Zeilstra, M. P. and Van Der Weide, R. (2013) 'Humans as an asset in a system consideration on the contribution of humans to system performance and system safety', *4th International Conference on Rail Human Factors* London. CRC Press, pp. 473-482.

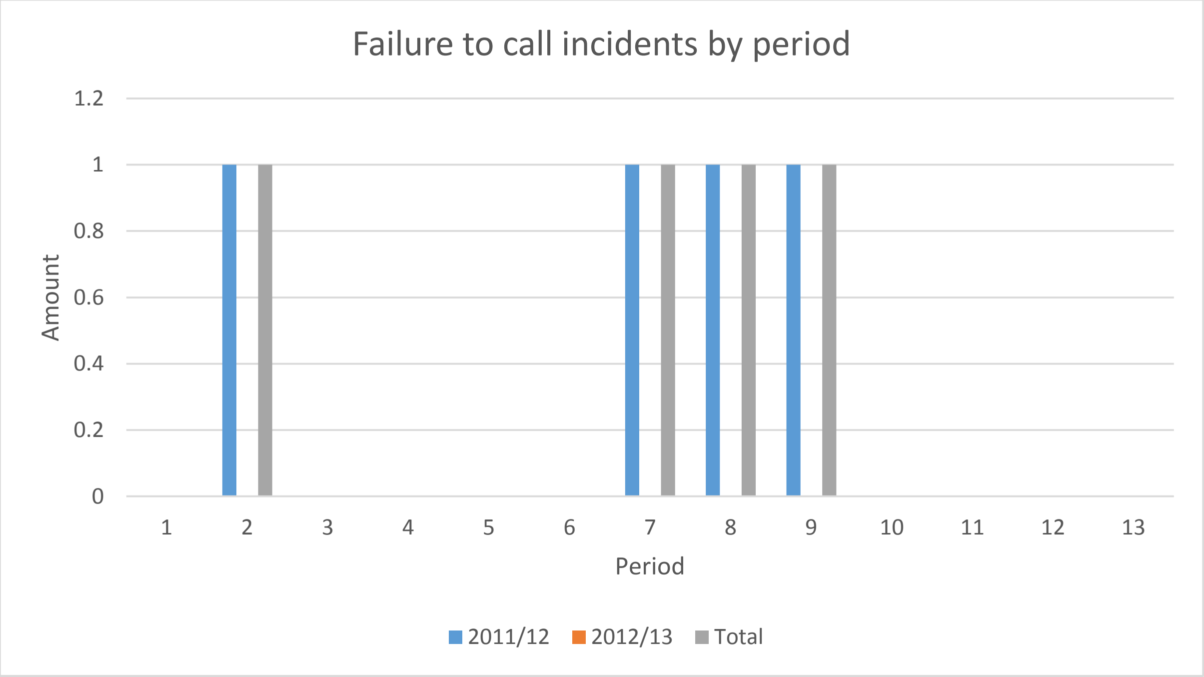
Zhang, J., Patel, V. L., Johnson, T. R. and Shortliffe, E. H. (2004) 'A cognitive taxonomy of medical errors', *Journal of Biomedical Informatics*, 37(3), pp. 193-204.

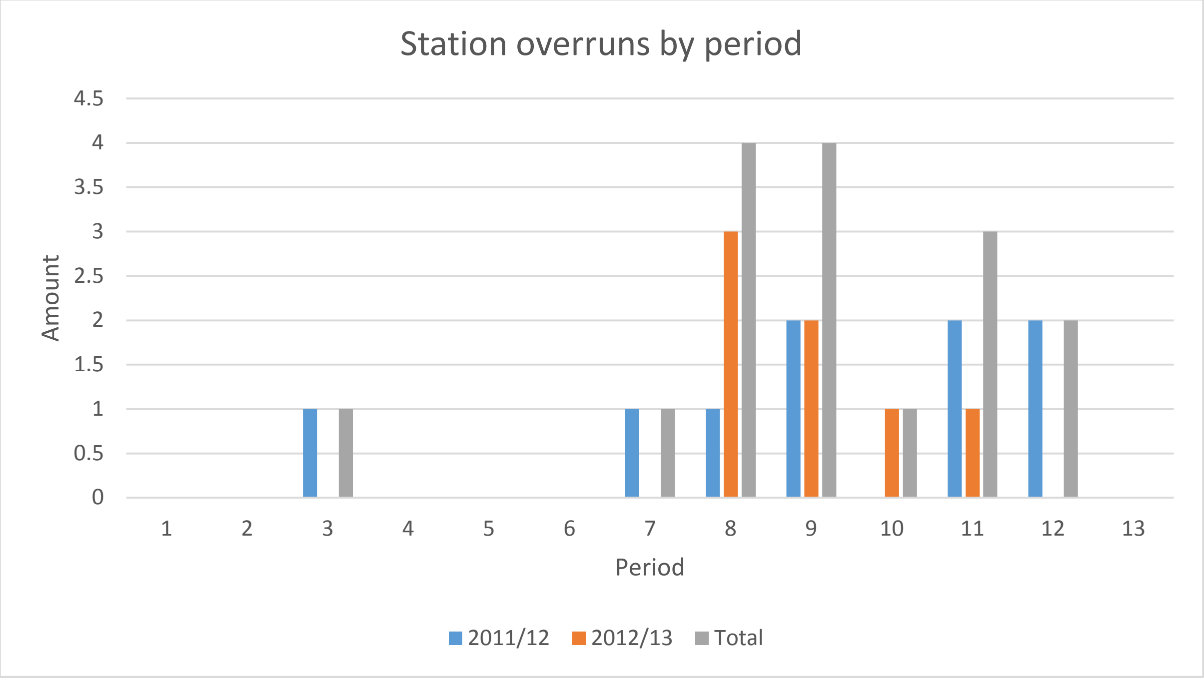
Zhu, Y., Olofsson, U. and Persson, K. (2012) 'Investigation of factors influencing wheel-rail adhesion using a mini-traction machine', *Wear*, 292-293, pp. 218-231.

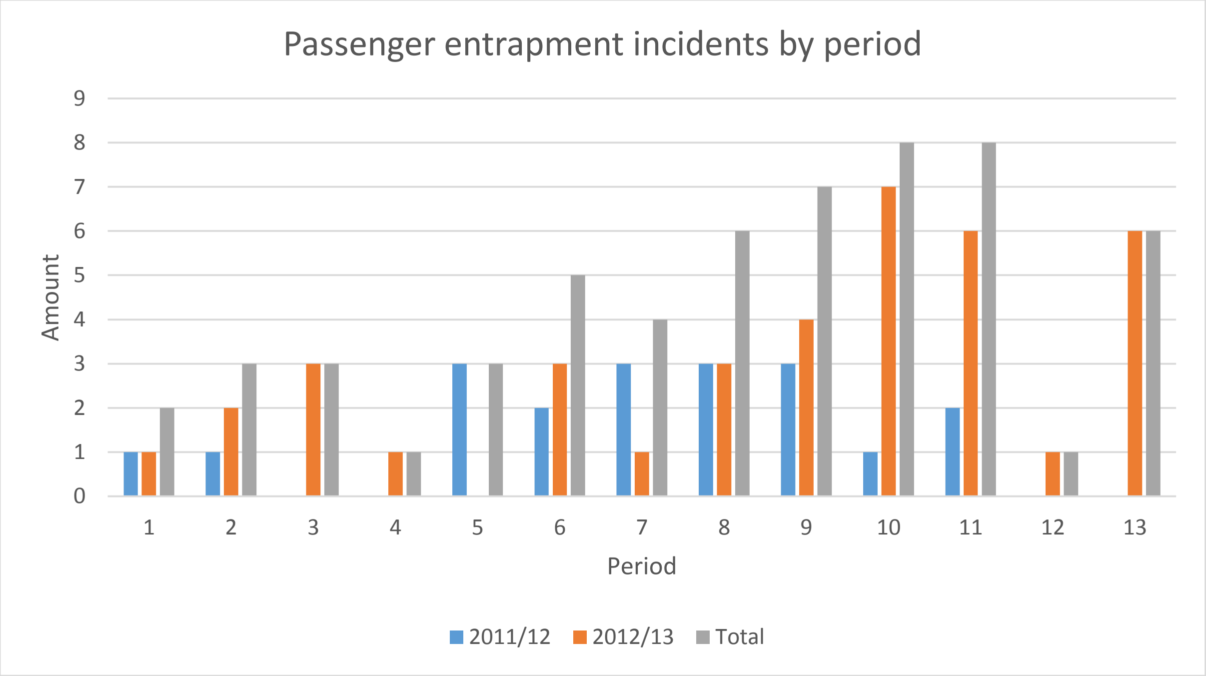
# Appendix A. Past driver-related incidents statistics by period

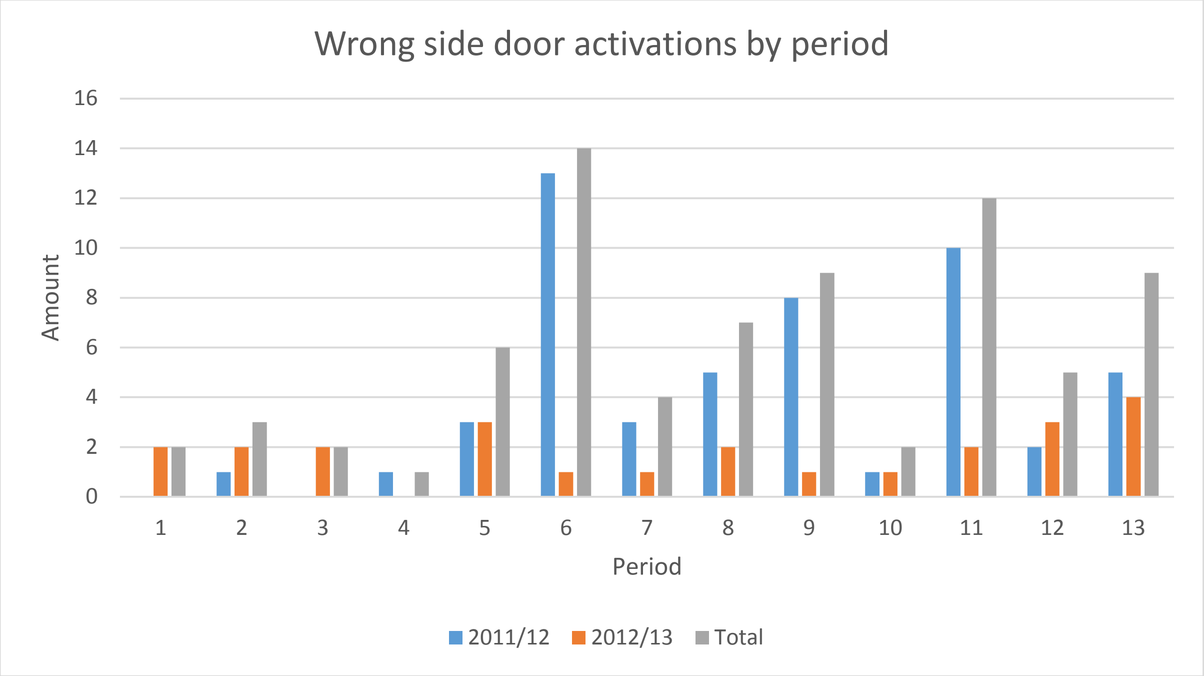


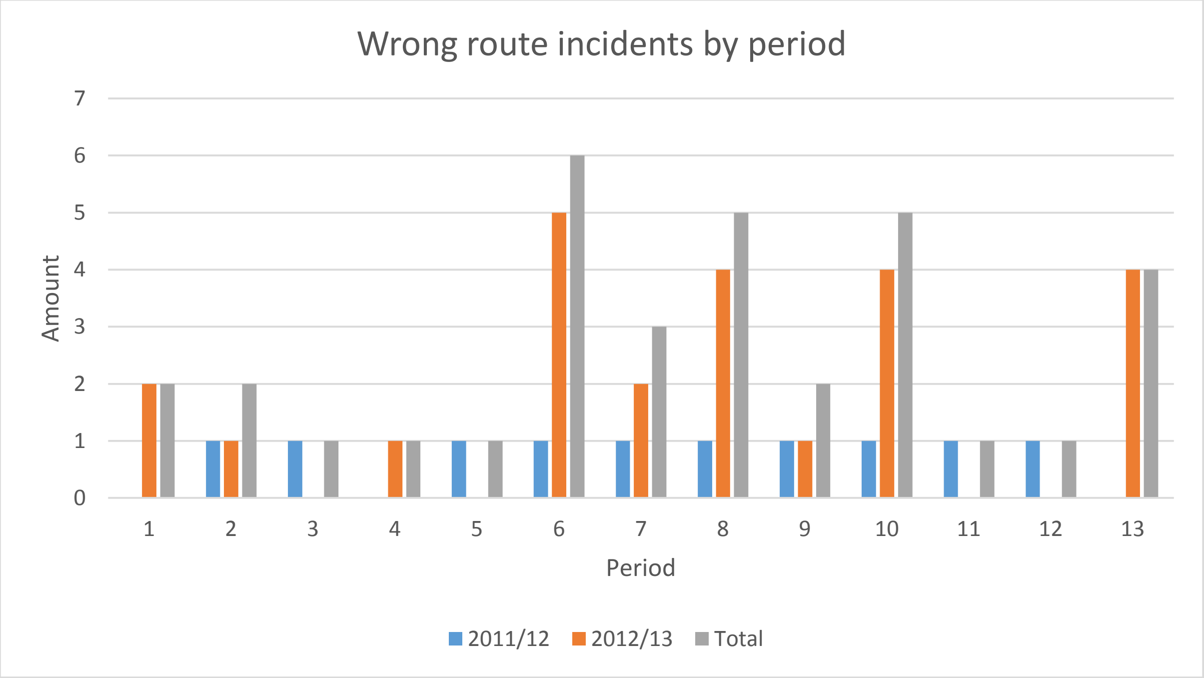



Failure to call incidents by period
Amount
1.2
1
0.8
0.6
0.4
0.2
0
1
2
3
4
5
6
7
8
9
10
11
12
13
Period
2011/12
2012/13
Total

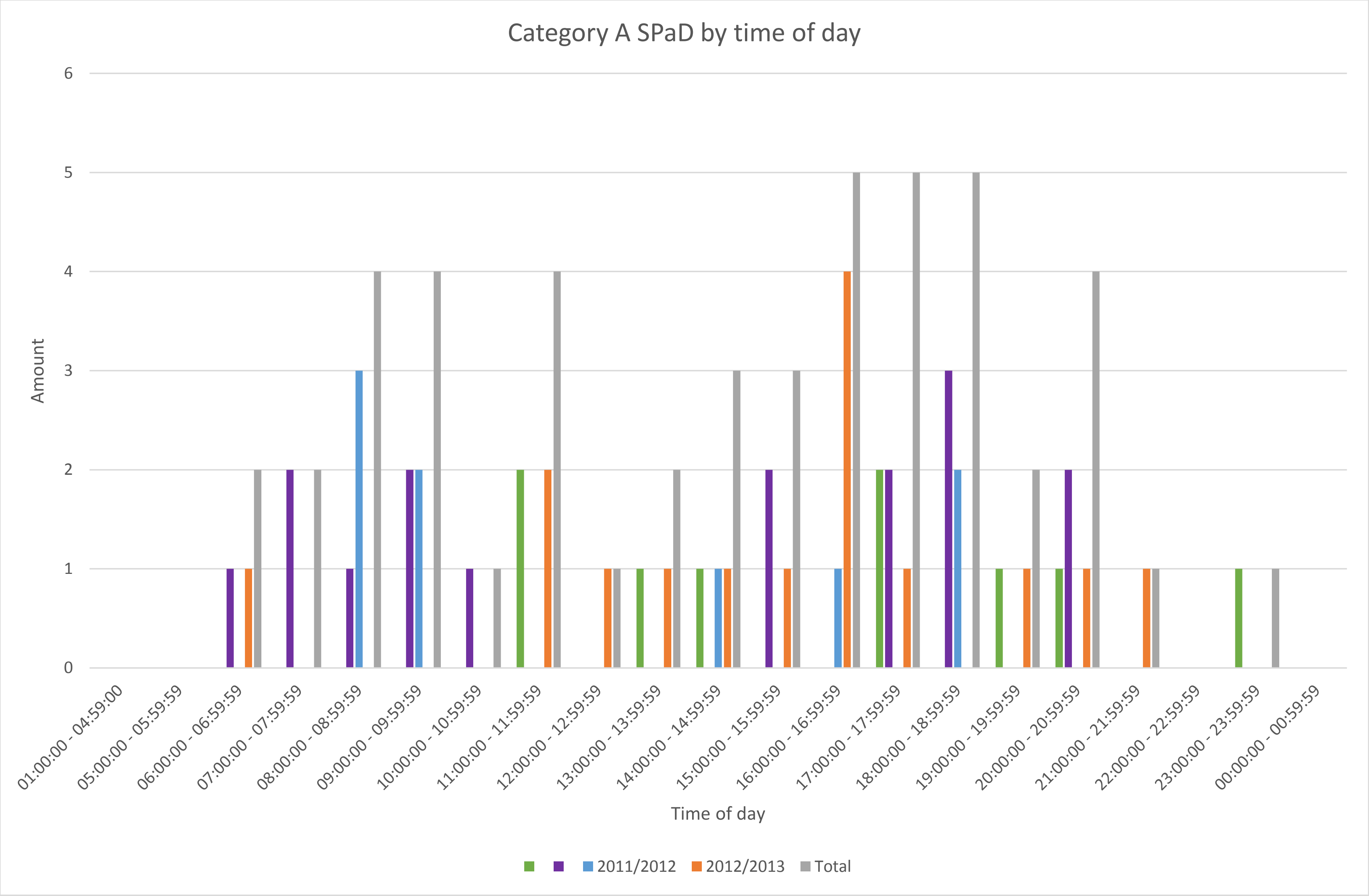

Station overruns by period
Amount
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
1
2
3
4
5
6
7
8
9
10
11
12
13
Period
2011/12
2012/13
Total

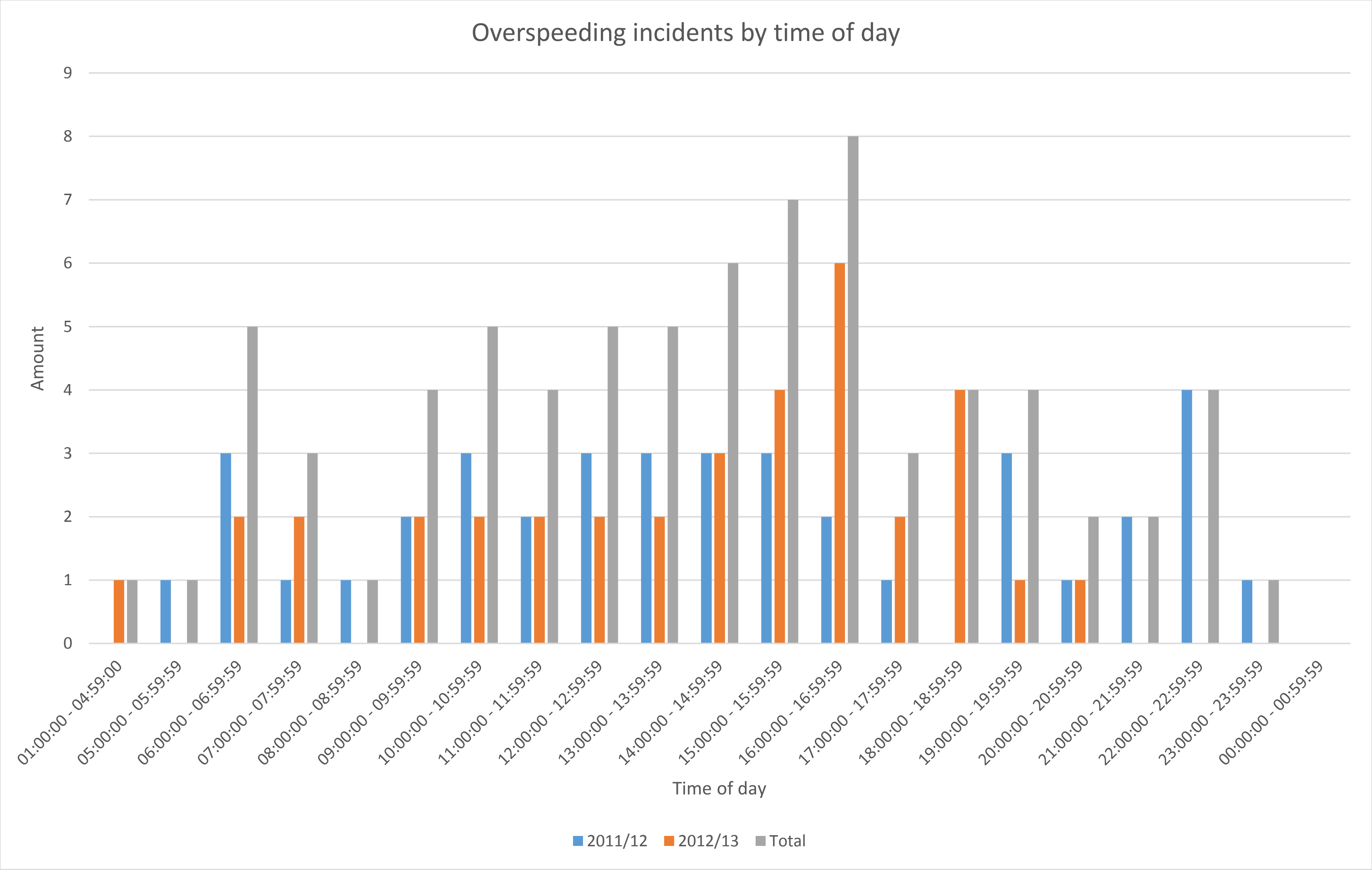

Passenger entrapment incidents by period
Amount
0
1
2
3
4
5
6
7
8
9
1
2
3
4
5
6
7
8
9
10
11
12
13
Period
2011/12
2012/13
Total

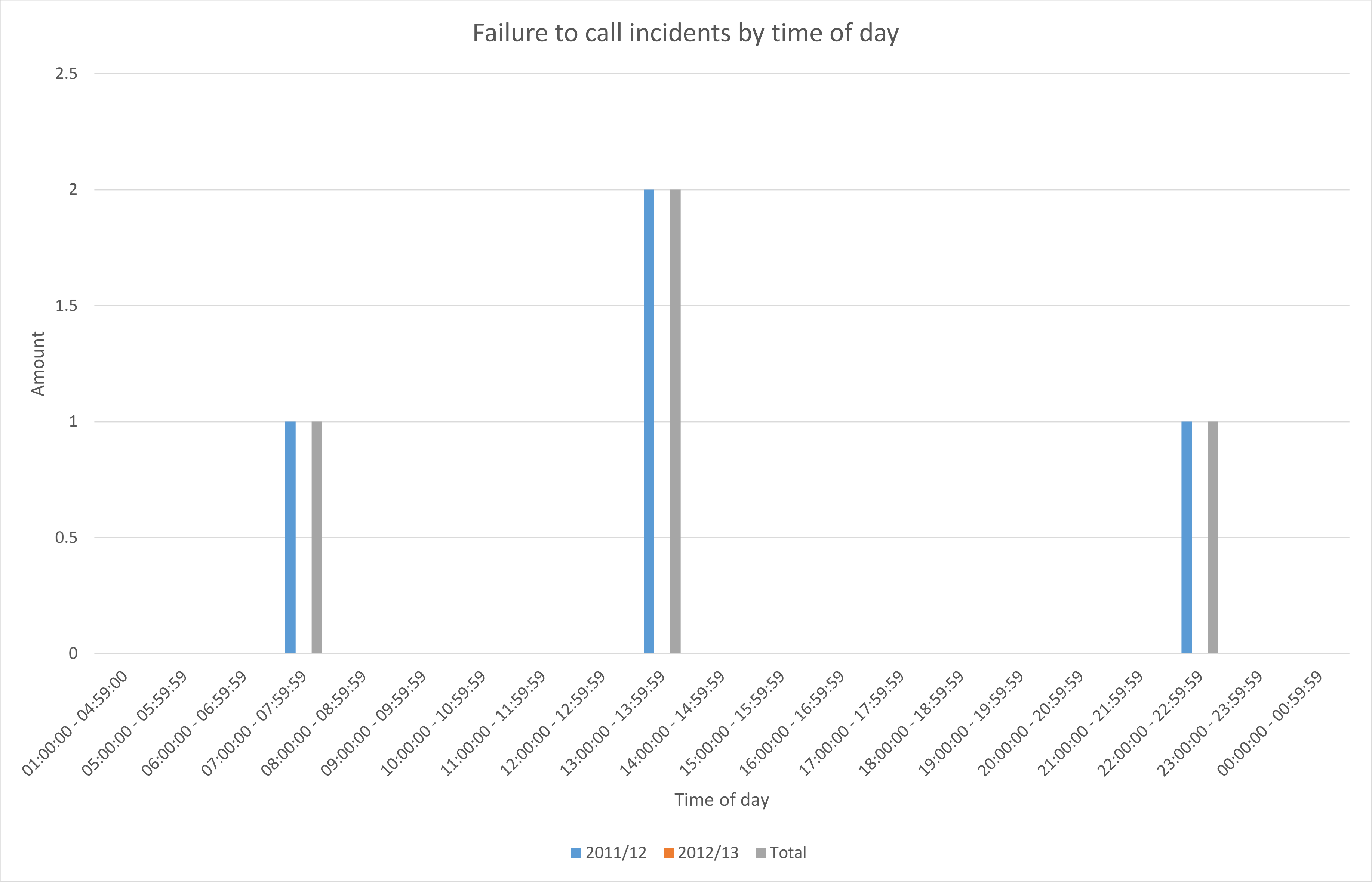

Wrong side door activations by period
Amount
0
2
4
6
8
10
12
14
16
1
2
3
4
5
6
7
8
9
10
11
12
13
Period
2011/12
2012/13
Total

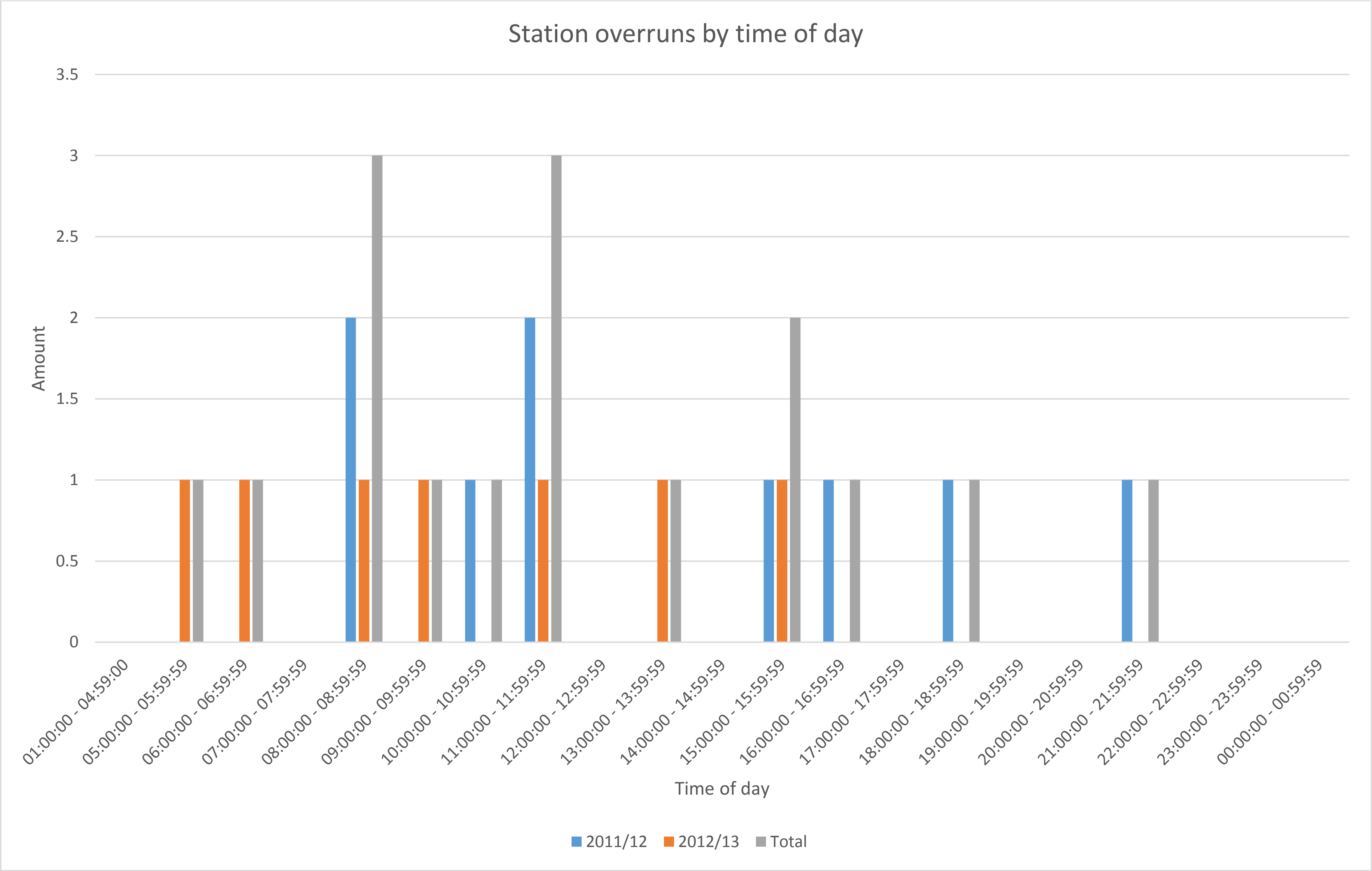

Wrong route incidents by period
Amount
0
1
2
3
4
5
6
7
Period
1
2
3
4
5
6
7
8
9
10
11
12
13
2011/12
2012/13
Total

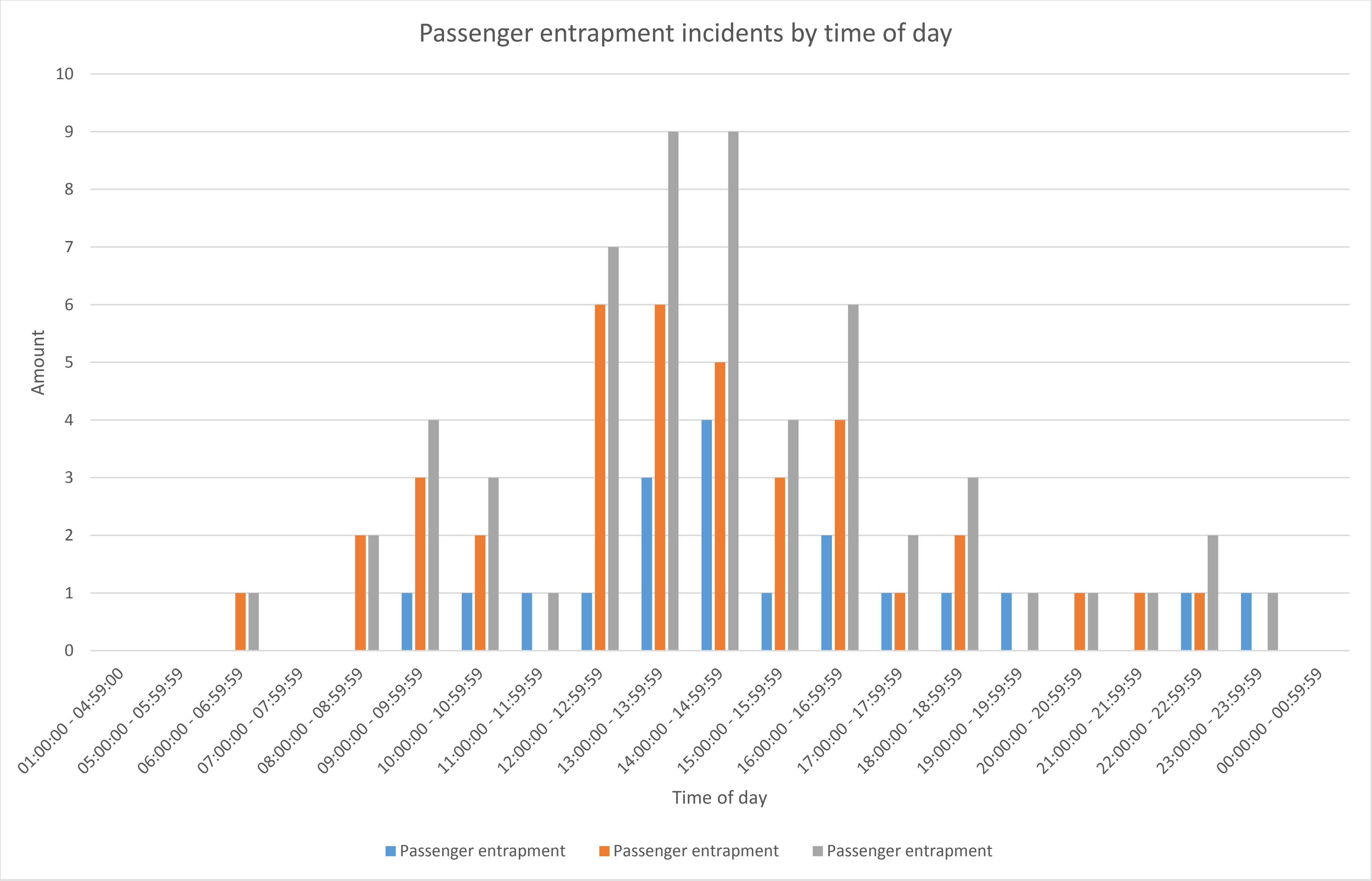
# Appendix B. Past driver-related incidents statistics by time of a day

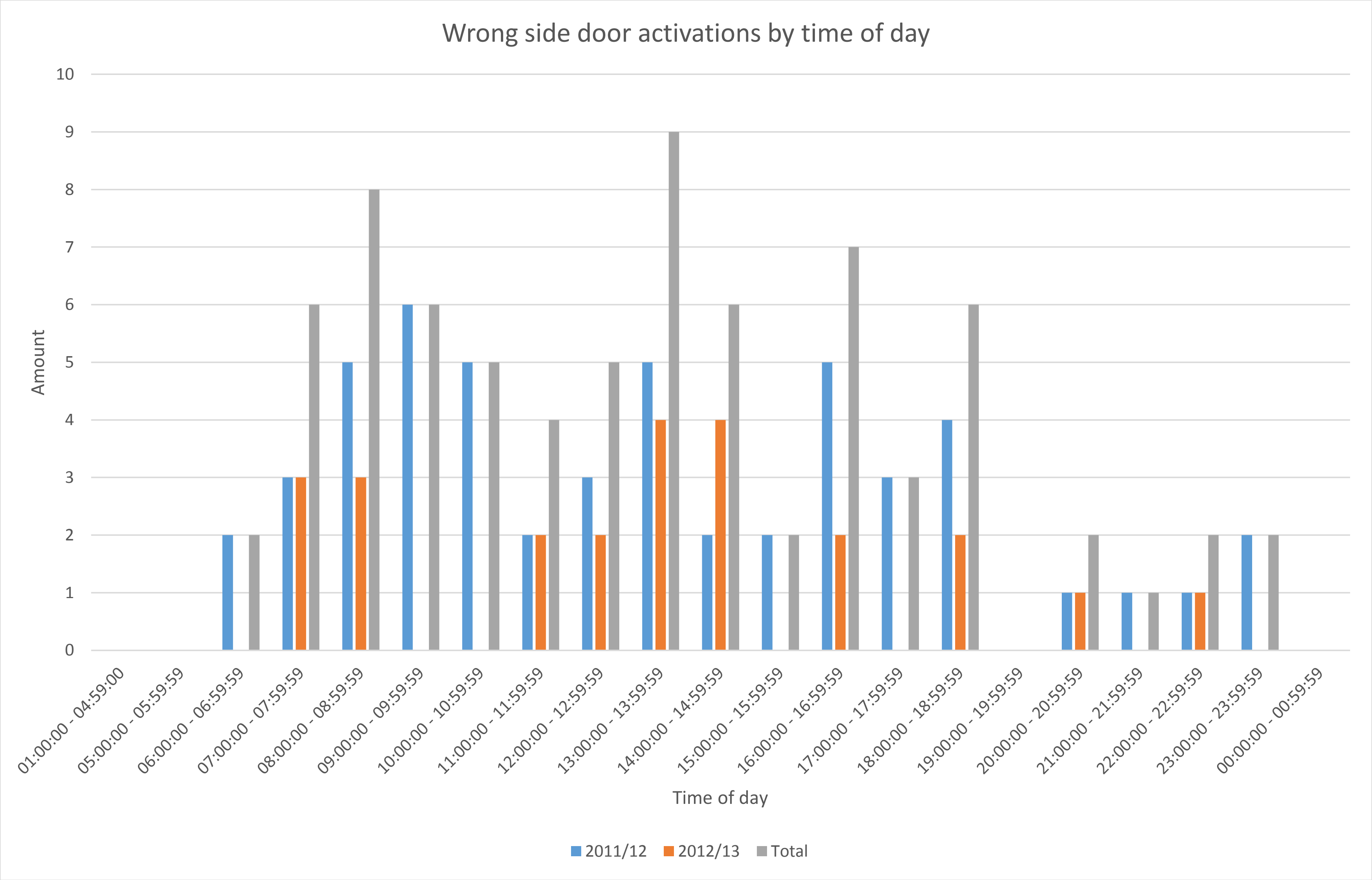

Category A SPaD by time of day
Amount
6
5
4
3
2
1
0
01:00:00 - 04:59:00
05:00:00 - 05:59:59
06:00:00 - 06:59:59
07:00:00 - 07:59:59
08:00:00 - 08:59:59
09:00:00 - 09:59:59
10:00:00 - 10:59:59
11:00:00 - 11:59:59
12:00:00 - 12:59:59
13:00:00 - 13:59:59
14:00:00 - 14:59:59
15:00:00 - 15:59:59
16:00:00 - 16:59:59
17:00:00 - 17:59:59
18:00:00 - 18:59:59
19:00:00 - 19:59:59
20:00:00 - 20:59:59
21:00:00 - 21:59:59
22:00:00 - 22:59:59
23:00:00 - 23:59:59
00:00:00 - 00:59:59
Time of day
🟩 🟪 🟦 2011/2012 🟧 2012/2013 ⬜ Total

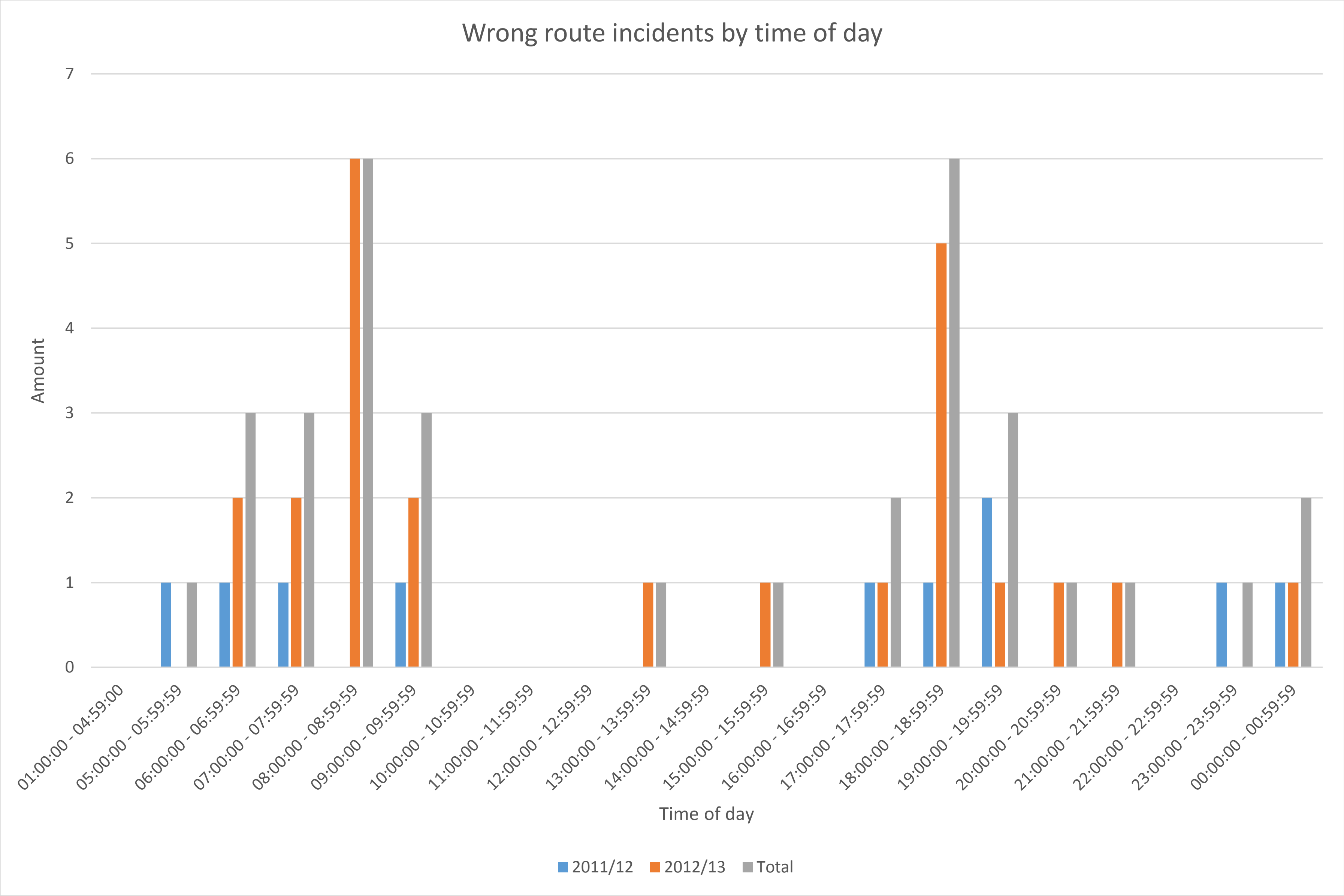

Overspeeding incidents by time of day
Amount
0
1
2
3
4
5
6
7
8
9
01:00:00 - 04:59:00
05:00:00 - 05:59:59
06:00:00 - 06:59:59
07:00:00 - 07:59:59
08:00:00 - 08:59:59
09:00:00 - 09:59:59
10:00:00 - 10:59:59
11:00:00 - 11:59:59
12:00:00 - 12:59:59
13:00:00 - 13:59:59
14:00:00 - 14:59:59
15:00:00 - 15:59:59
16:00:00 - 16:59:59
17:00:00 - 17:59:59
18:00:00 - 18:59:59
19:00:00 - 19:59:59
20:00:00 - 20:59:59
21:00:00 - 21:59:59
22:00:00 - 22:59:59
23:00:00 - 23:59:59
00:00:00 - 00:59:59
Time of day
2011/12
2012/13
Total


Failure to call incidents by time of day
2.5
2
1.5
1
0.5
0
Amount
01:00:00 - 04:59:00
05:00:00 - 05:59:59
06:00:00 - 06:59:59
07:00:00 - 07:59:59
08:00:00 - 08:59:59
09:00:00 - 09:59:59
10:00:00 - 10:59:59
11:00:00 - 11:59:59
12:00:00 - 12:59:59
13:00:00 - 13:59:59
14:00:00 - 14:59:59
15:00:00 - 15:59:59
16:00:00 - 16:59:59
17:00:00 - 17:59:59
18:00:00 - 18:59:59
19:00:00 - 19:59:59
20:00:00 - 20:59:59
21:00:00 - 21:59:59
22:00:00 - 22:59:59
23:00:00 - 23:59:59
00:00:00 - 00:59:59
Time of day
2011/12
2012/13
Total

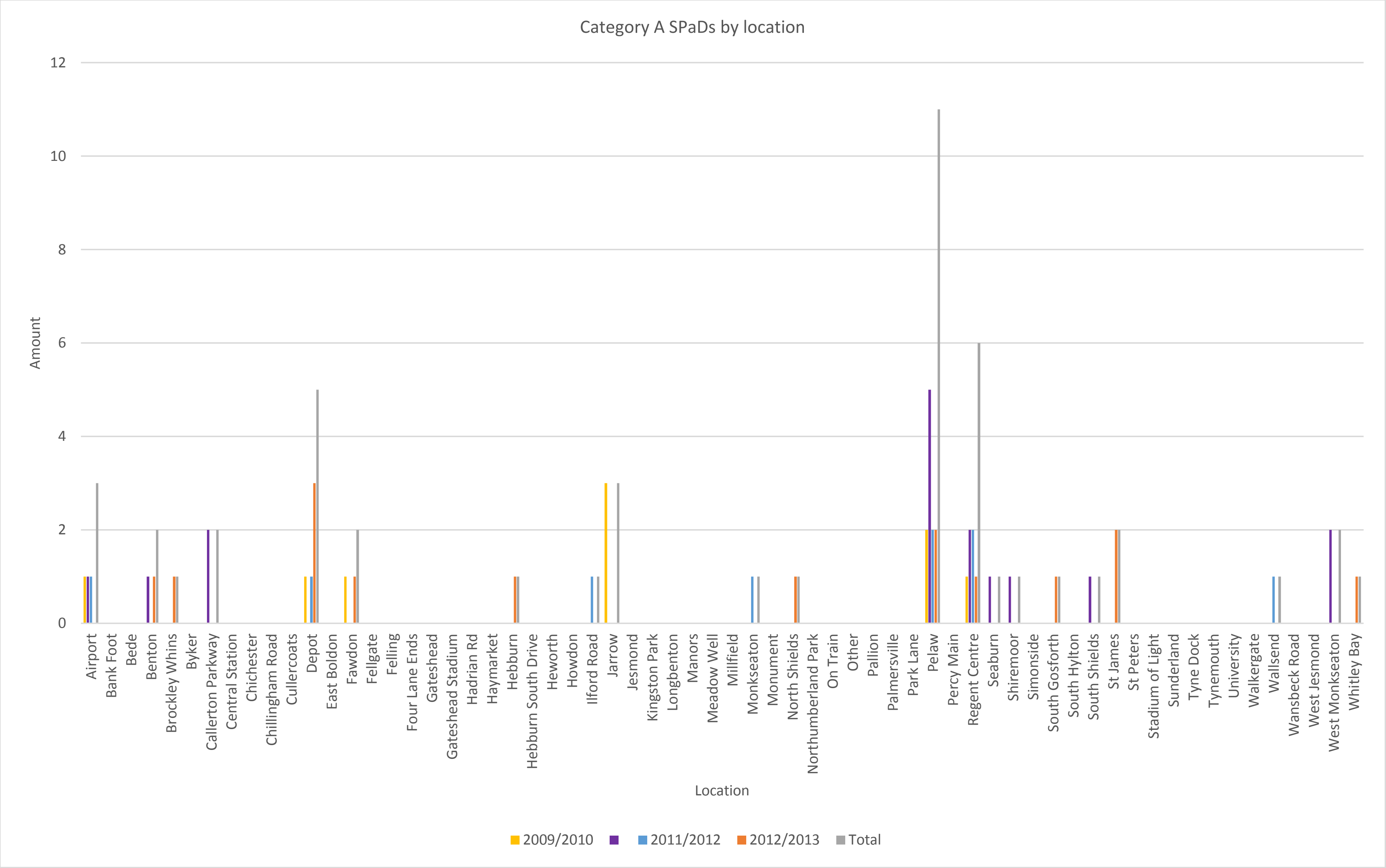

Station overruns by time of day
Amount
3.5
3
2.5
2
1.5
1
0.5
0
01:00:00 - 04:59:00
05:00:00 - 05:59:59
06:00:00 - 06:59:59
07:00:00 - 07:59:59
08:00:00 - 08:59:59
09:00:00 - 09:59:59
10:00:00 - 10:59:59
11:00:00 - 11:59:59
12:00:00 - 12:59:59
13:00:00 - 13:59:59
14:00:00 - 14:59:59
15:00:00 - 15:59:59
16:00:00 - 16:59:59
17:00:00 - 17:59:59
18:00:00 - 18:59:59
19:00:00 - 19:59:59
20:00:00 - 20:59:59
21:00:00 - 21:59:59
22:00:00 - 22:59:59
23:00:00 - 23:59:59
00:00:00 - 00:59:59
Time of day
2011/12
2012/13
Total

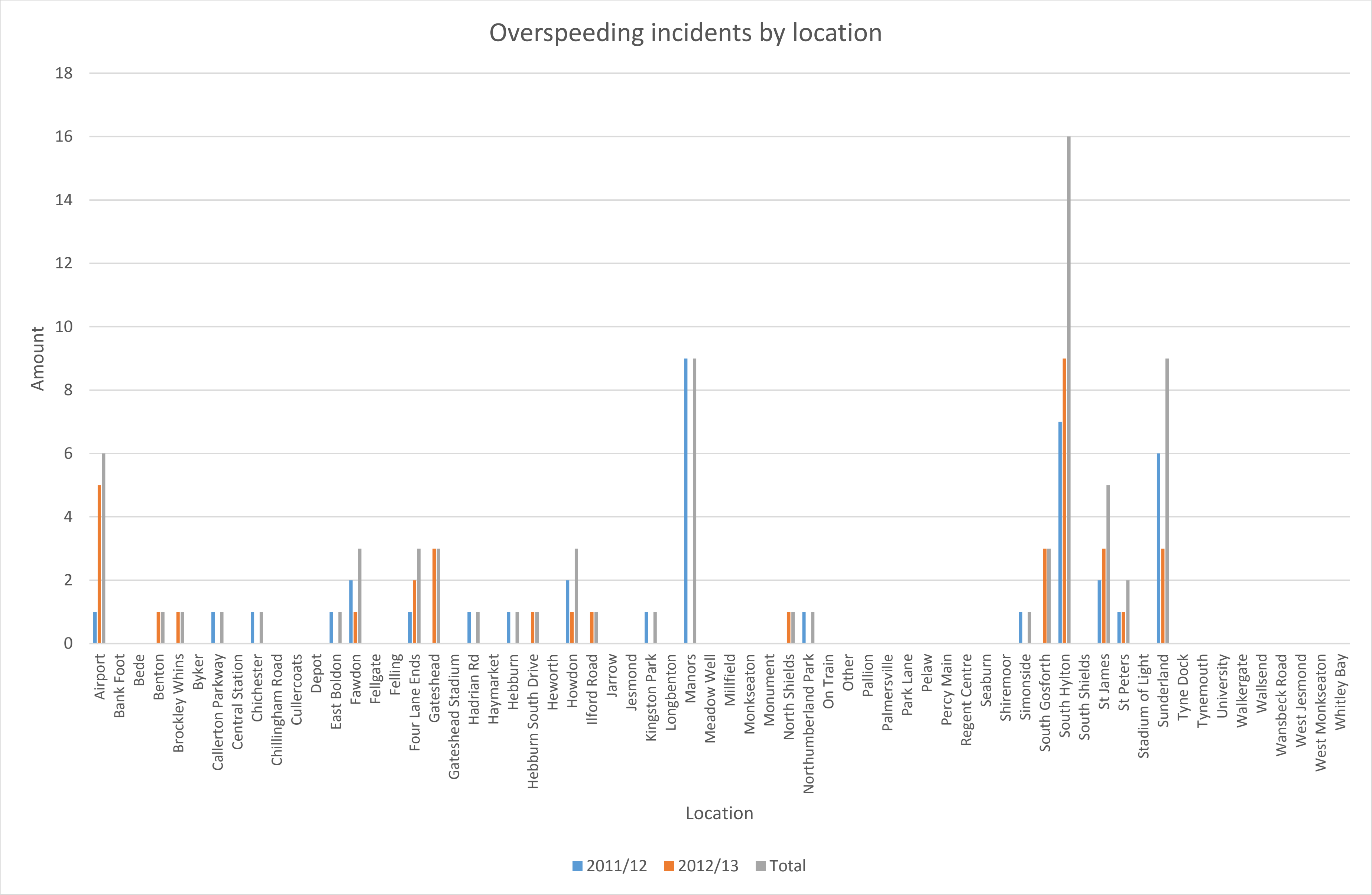

Passenger entrapment incidents by time of day
10
9
8
7
6
5
4
3
2
1
0
Amount
01:00:00 - 04:59:00
05:00:00 - 05:59:59
06:00:00 - 06:59:59
07:00:00 - 07:59:59
08:00:00 - 08:59:59
09:00:00 - 09:59:59
10:00:00 - 10:59:59
11:00:00 - 11:59:59
12:00:00 - 12:59:59
13:00:00 - 13:59:59
14:00:00 - 14:59:59
15:00:00 - 15:59:59
16:00:00 - 16:59:59
17:00:00 - 17:59:59
18:00:00 - 18:59:59
19:00:00 - 19:59:59
20:00:00 - 20:59:59
21:00:00 - 21:59:59
22:00:00 - 22:59:59
23:00:00 - 23:59:59
00:00:00 - 00:59:59
Time of day
Passenger entrapment
Passenger entrapment
Passenger entrapment

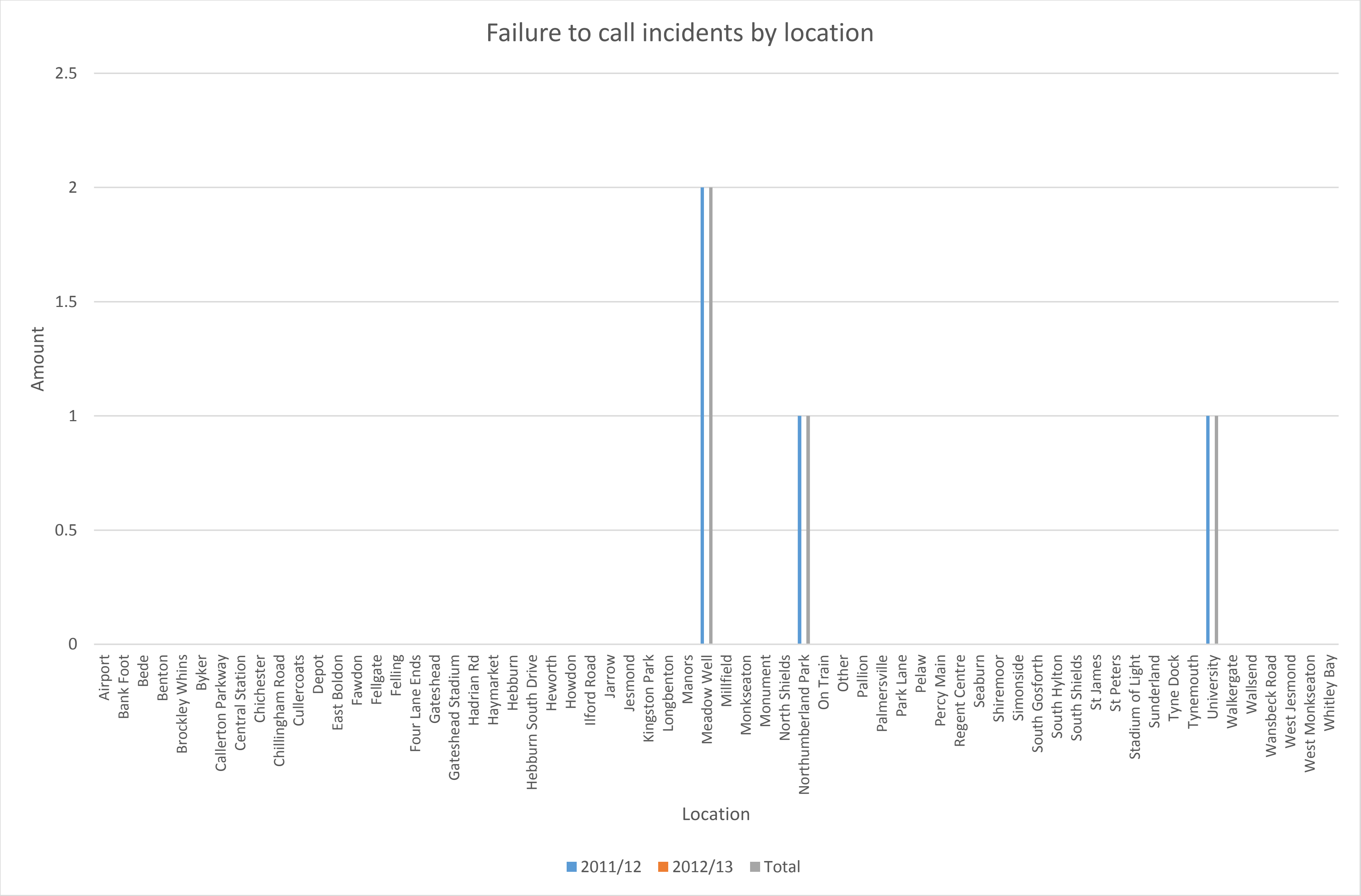

Wrong side door activations by time of day
Amount
0
1
2
3
4
5
6
7
8
9
10
01:00:00 - 04:59:00
05:00:00 - 05:59:59
06:00:00 - 06:59:59
07:00:00 - 07:59:59
08:00:00 - 08:59:59
09:00:00 - 09:59:59
10:00:00 - 10:59:59
11:00:00 - 11:59:59
12:00:00 - 12:59:59
13:00:00 - 13:59:59
14:00:00 - 14:59:59
15:00:00 - 15:59:59
16:00:00 - 16:59:59
17:00:00 - 17:59:59
18:00:00 - 18:59:59
19:00:00 - 19:59:59
20:00:00 - 20:59:59
21:00:00 - 21:59:59
22:00:00 - 22:59:59
23:00:00 - 23:59:59
00:00:00 - 00:59:59
Time of day
2011/12
2012/13
Total

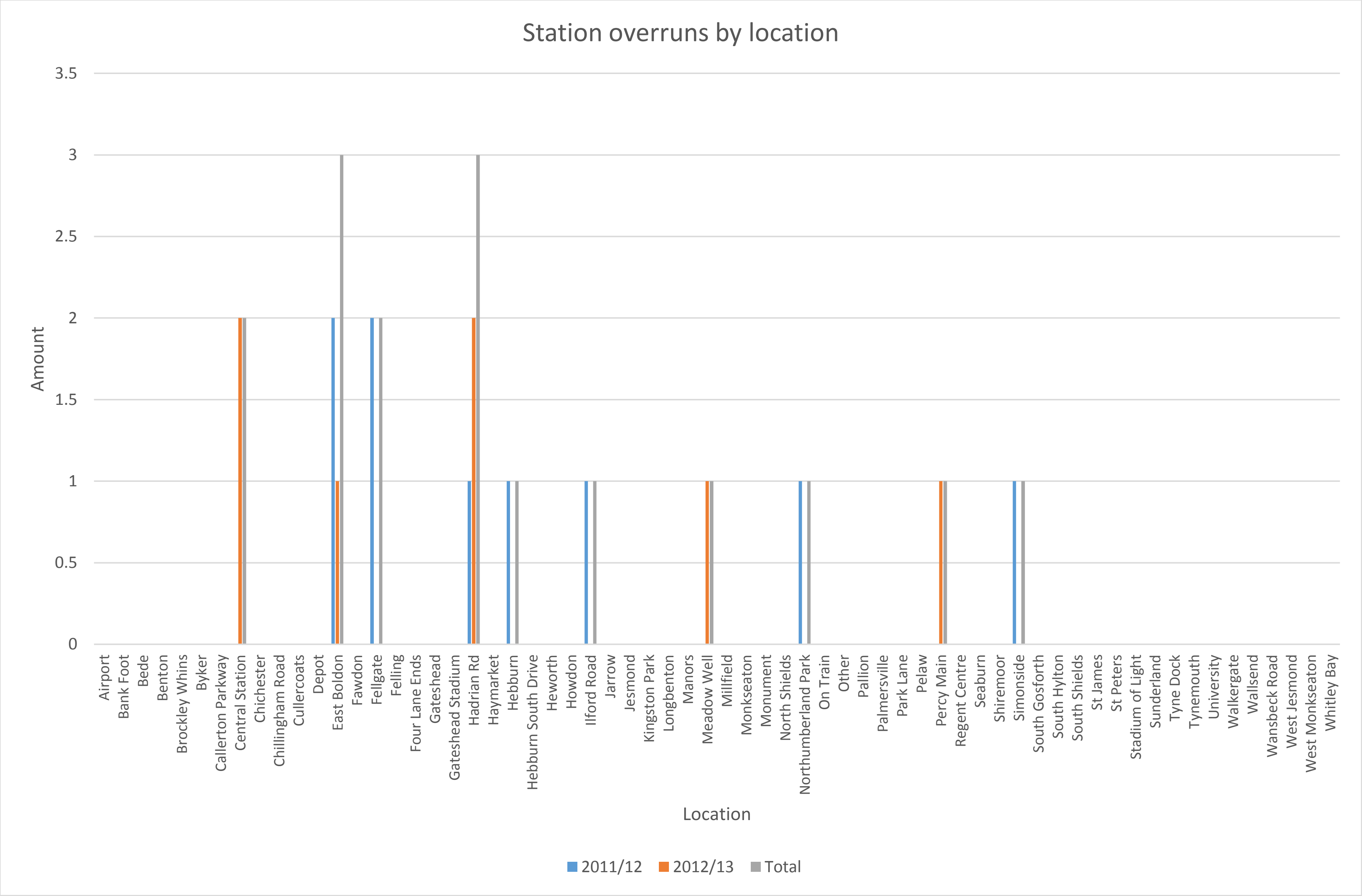

Wrong route incidents by time of day
Amount
0
1
2
3
4
5
6
7
Time of day
01:00:00 - 04:59:00
05:00:00 - 05:59:59
06:00:00 - 06:59:59
07:00:00 - 07:59:59
08:00:00 - 08:59:59
09:00:00 - 09:59:59
10:00:00 - 10:59:59
11:00:00 - 11:59:59
12:00:00 - 12:59:59
13:00:00 - 13:59:59
14:00:00 - 14:59:59
15:00:00 - 15:59:59
16:00:00 - 16:59:59
17:00:00 - 17:59:59
18:00:00 - 18:59:59
19:00:00 - 19:59:59
20:00:00 - 20:59:59
21:00:00 - 21:59:59
22:00:00 - 22:59:59
23:00:00 - 23:59:59
00:00:00 - 00:59:59
2011/12
2012/13
Total

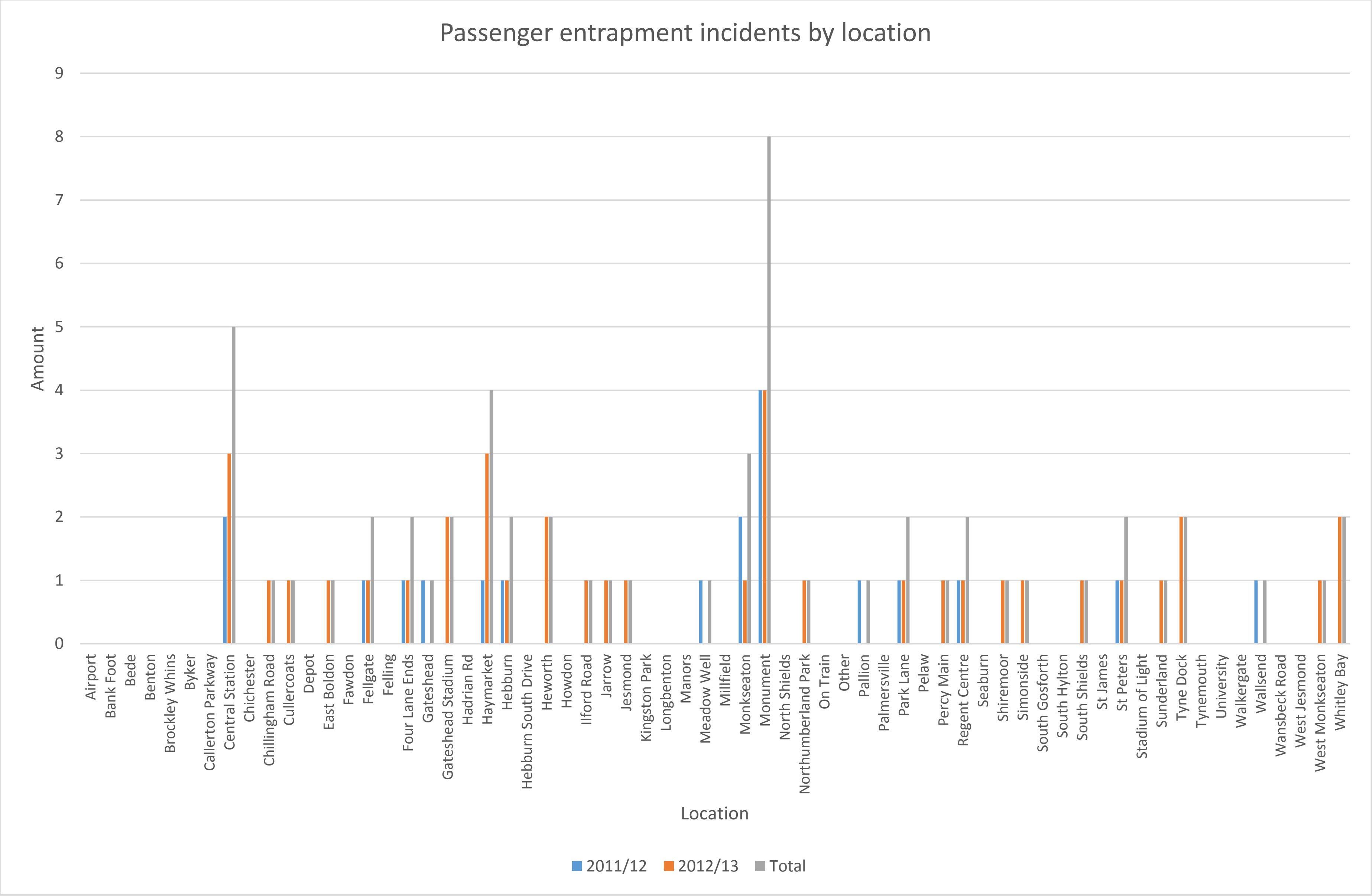
# Appendix C. Past driver-related incidents statistics by location

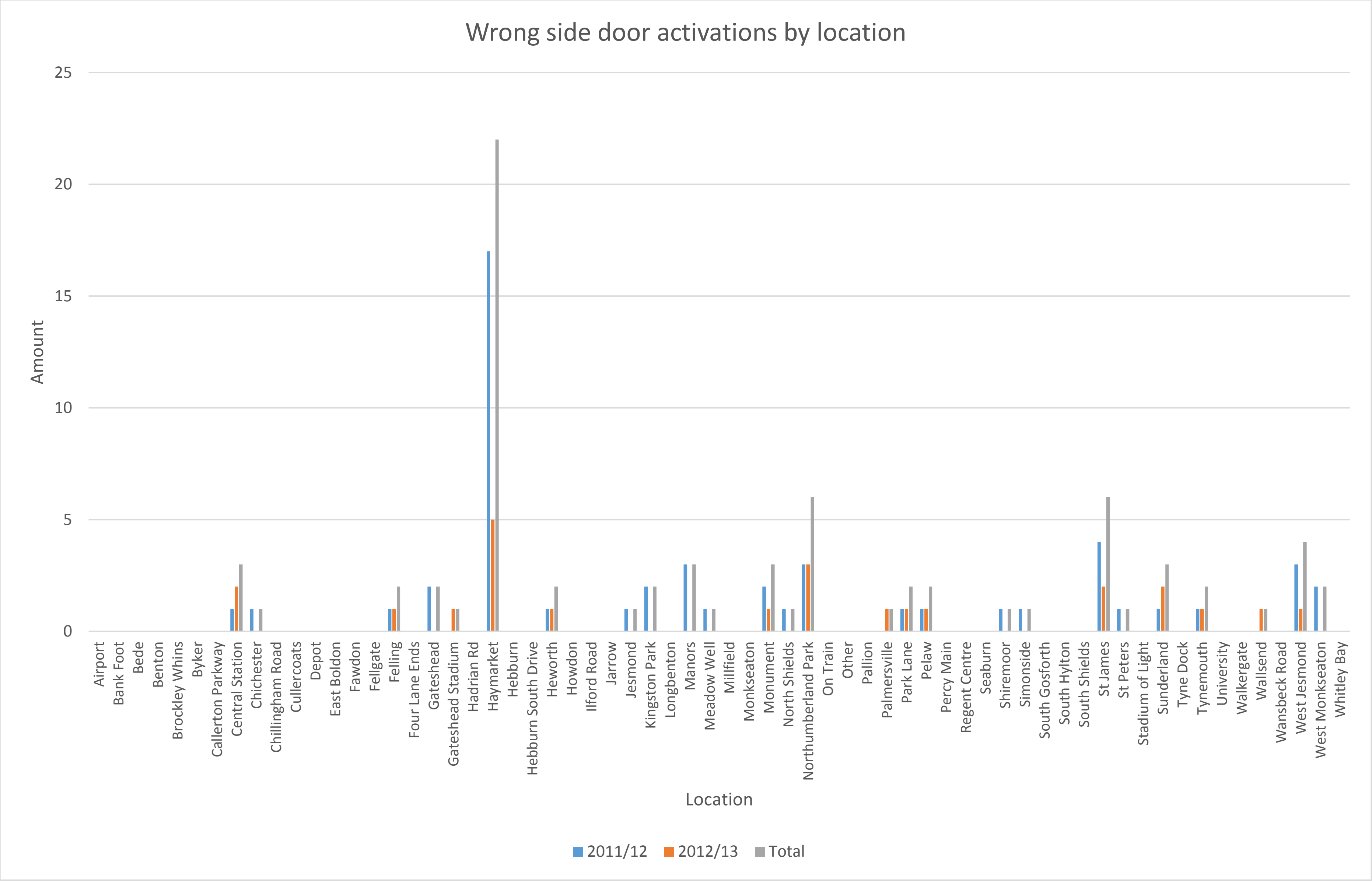

Category A SPaDs by location
12
10
8
6
4
2
0
Amount
Airport
Bank Foot
Bede
Benton
Brockley Whins
Byker
Callerton Parkway
Central Station
Chichester
Chillingham Road
Cullercoats
Depot
East Boldon
Fawdon
Fellgate
Felling
Four Lane Ends
Gateshead
Gateshead Stadium
Hadrian Rd
Haymarket
Hebburn
Hebburn South Drive
Heworth
Howdon
Ilford Road
Jarrow
Jesmond
Kingston Park
Longbenton
Manors
Meadow Well
Millfield
Monkseaton
Monument
North Shields
Northumberland Park
On Train
Other
Pallion
Palmersville
Park Lane
Pelaw
Percy Main
Regent Centre
Seaburn
Shiremoor
Simonside
South Gosforth
South Hylton
South Shields
St James
St Peters
Stadium of Light
Sunderland
Tyne Dock
Tynemouth
University
Walkergate
Wallsend
Wansbeck Road
West Jesmond
West Monkseaton
Whitley Bay
Location
2009/2010
2011/2012
2012/2013
Total

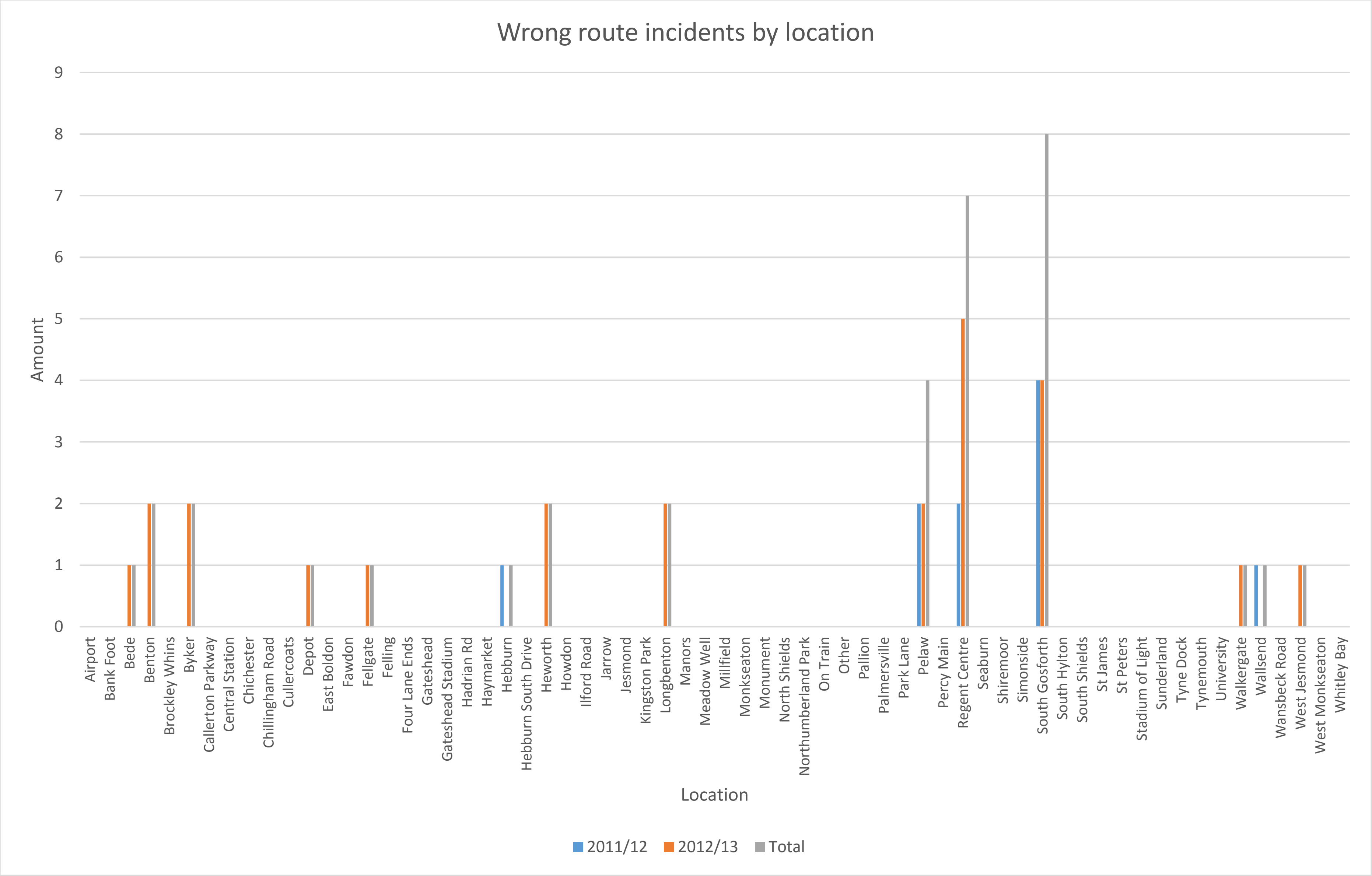

Overspeeding incidents by location
Amount
18
16
14
12
10
8
6
4
2
0
Airport
Bank Foot
Bede
Benton
Brockley Whins
Byker
Callerton Parkway
Central Station
Chichester
Chillingham Road
Cullercoats
Depot
East Boldon
Fawdon
Fellgate
Felling
Four Lane Ends
Gateshead
Gateshead Stadium
Hadrian Rd
Haymarket
Hebburn
Hebburn South Drive
Heworth
Howdon
Ilford Road
Jarrow
Jesmond
Kingston Park
Longbenton
Manors
Meadow Well
Millfield
Monkseaton
Monument
North Shields
Northumberland Park
On Train
Other
Pallion
Palmersville
Park Lane
Pelaw
Percy Main
Regent Centre
Seaburn
Shiremoor
Simonside
South Gosforth
South Hylton
South Shields
St James
St Peters
Stadium of Light
Sunderland
Tyne Dock
Tynemouth
University
Walkergate
Wallsend
Wansbeck Road
West Jesmond
West Monkseaton
Whitley Bay
Location
2011/12
2012/13
Total


Failure to call incidents by location
Amount
2.5
2
1.5
1
0.5
0
Location
Airport
Bank Foot
Bede
Benton
Brockley Whins
Byker
Callerton Parkway
Central Station
Chichester
Chillingham Road
Cullercoats
Depot
East Boldon
Fawdon
Fellgate
Felling
Four Lane Ends
Gateshead
Gateshead Stadium
Hadrian Rd
Haymarket
Hebburn
Hebburn South Drive
Heworth
Howdon
Ilford Road
Jarrow
Jesmond
Kingston Park
Longbenton
Manors
Meadow Well
Millfield
Monkseaton
Monument
North Shields
Northumberland Park
On Train
Other
Pallion
Palmersville
Park Lane
Pelaw
Percy Main
Regent Centre
Seaburn
Shiremoor
Simonside
South Gosforth
South Hylton
South Shields
St James
St Peters
Stadium of Light
Sunderland
Tyne Dock
Tynemouth
University
Walkergate
Wallsend
Wansbeck Road
West Jesmond
West Monkseaton
Whitley Bay
2011/12
2012/13
Total

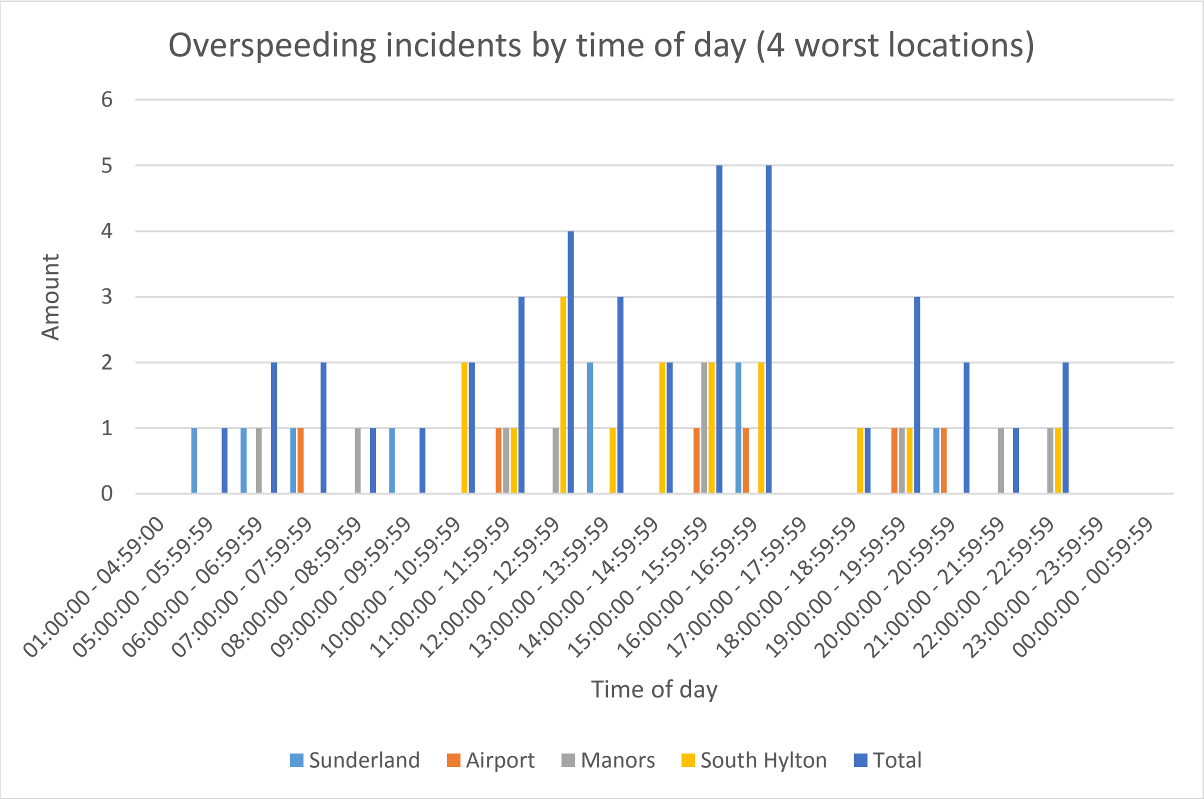

Station overruns by location
Amount
3.5
3
2.5
2
1.5
1
0.5
0
Airport
Bank Foot
Bede
Benton
Brockley Whins
Byker
Callerton Parkway
Central Station
Chichester
Chillingham Road
Cullercoats
Depot
East Boldon
Fawdon
Fellgate
Felling
Four Lane Ends
Gateshead
Gateshead Stadium
Hadrian Rd
Haymarket
Hebburn
Hebburn South Drive
Heworth
Howdon
Ilford Road
Jarrow
Jesmond
Kingston Park
Longbenton
Manors
Meadow Well
Millfield
Monkseaton
Monument
North Shields
Northumberland Park
On Train
Other
Pallion
Palmersville
Park Lane
Pelaw
Percy Main
Regent Centre
Seaburn
Shiremoor
Simonside
South Gosforth
South Hylton
South Shields
St James
St Peters
Stadium of Light
Sunderland
Tyne Dock
Tynemouth
University
Walkergate
Wallsend
Wansbeck Road
West Jesmond
West Monkseaton
Whitley Bay
Location
■ 2011/12
■ 2012/13
■ Total

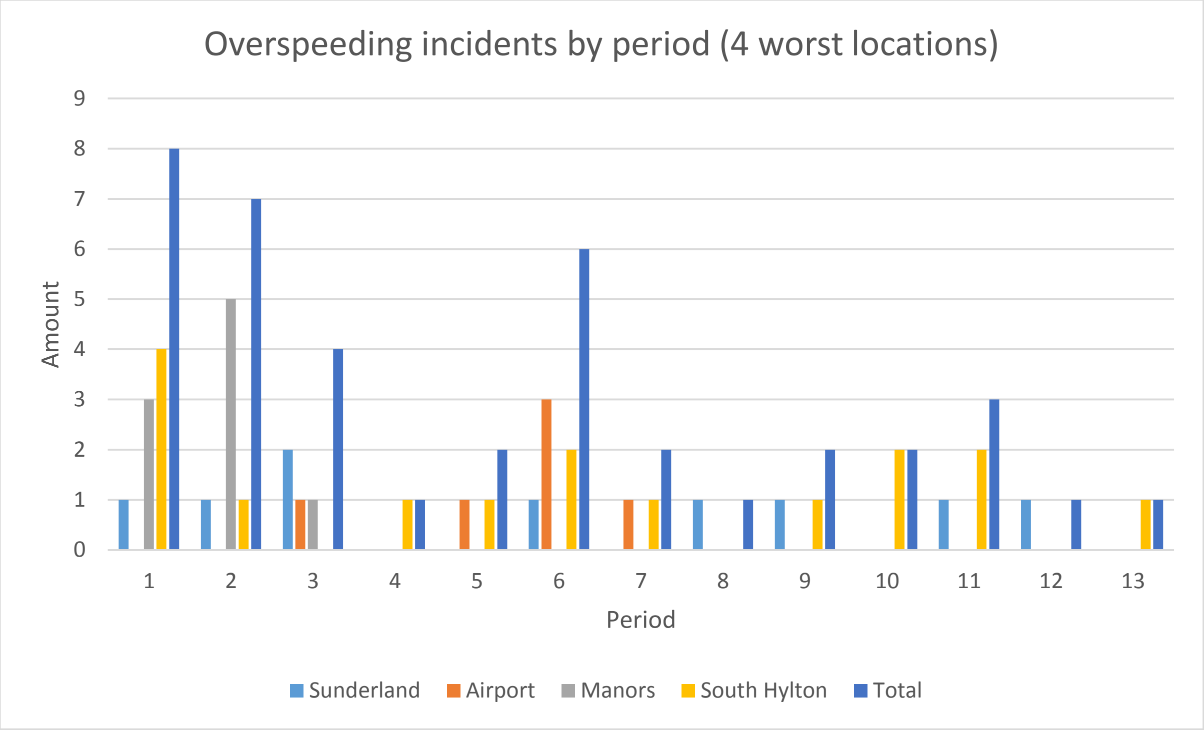

Passenger entrapment incidents by location
Amount
0
1
2
3
4
5
6
7
8
9
Airport
Bank Foot
Bede
Benton
Brockley Whins
Byker
Callerton Parkway
Central Station
Chichester
Chillingham Road
Cullercoats
Depot
East Boldon
Fawdon
Fellgate
Felling
Four Lane Ends
Gateshead
Gateshead Stadium
Hadrian Rd
Haymarket
Hebburn
Hebburn South Drive
Heworth
Howdon
Ilford Road
Jarrow
Jesmond
Kingston Park
Longbenton
Manors
Meadow Well
Millfield
Monkseaton
Monument
North Shields
Northumberland Park
On Train
Other
Pallion
Palmersville
Park Lane
Pelaw
Percy Main
Regent Centre
Seaburn
Shiremoor
Simonside
South Gosforth
South Hylton
South Shields
St James
St Peters
Stadium of Light
Sunderland
Tyne Dock
Tynemouth
University
Walkergate
Wallsend
Wansbeck Road
West Jesmond
West Monkseaton
Whitley Bay
Location
■ 2011/12
■ 2012/13
■ Total


Wrong side door activations by location
25
20
15
10
5
0
Amount
Airport
Bank Foot
Bede
Benton
Brockley Whins
Byker
Callerton Parkway
Central Station
Chichester
Chillingham Road
Cullercoats
Depot
East Boldon
Fawdon
Fellgate
Felling
Four Lane Ends
Gateshead
Gateshead Stadium
Hadrian Rd
Haymarket
Hebburn
Hebburn South Drive
Heworth
Howdon
Ilford Road
Jarrow
Jesmond
Kingston Park
Longbenton
Manors
Meadow Well
Millfield
Monkseaton
Monument
North Shields
Northumberland Park
On Train
Other
Pallion
Palmersville
Park Lane
Pelaw
Percy Main
Regent Centre
Seaburn
Shiremoor
Simonside
South Gosforth
South Hylton
South Shields
St James
St Peters
Stadium of Light
Sunderland
Tyne Dock
Tynemouth
University
Walkergate
Wallsend
Wansbeck Road
West Jesmond
West Monkseaton
Whitley Bay
Location
2011/12
2012/13
Total

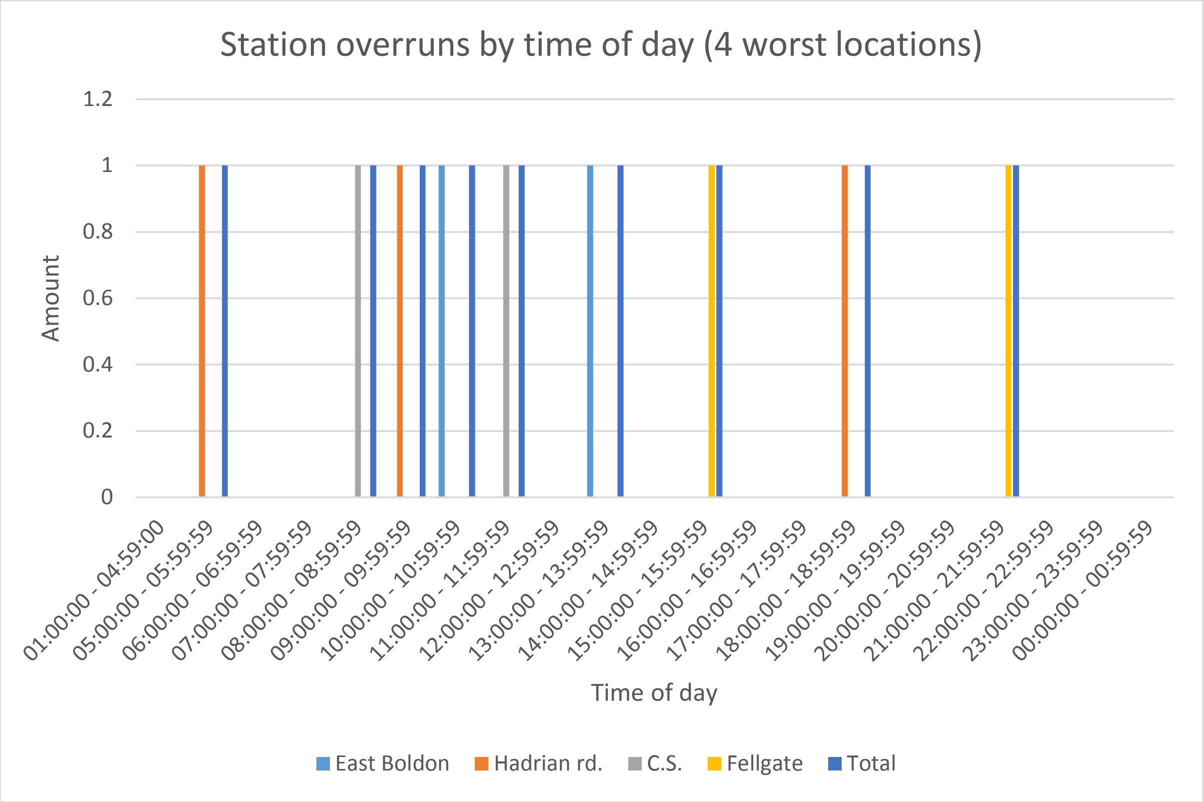

Wrong route incidents by location
Amount
9
8
7
6
5
4
3
2
1
0
Airport
Bank Foot
Bede
Benton
Brockley Whins
Byker
Callerton Parkway
Central Station
Chichester
Chillingham Road
Cullercoats
Depot
East Boldon
Fawdon
Fellgate
Felling
Four Lane Ends
Gateshead
Gateshead Stadium
Hadrian Rd
Haymarket
Hebburn
Hebburn South Drive
Heworth
Howdon
Ilford Road
Jarrow
Jesmond
Kingston Park
Longbenton
Manors
Meadow Well
Millfield
Monkseaton
Monument
North Shields
Northumberland Park
On Train
Other
Pallion
Palmersville
Park Lane
Pelaw
Percy Main
Regent Centre
Seaburn
Shiremoor
Simonside
South Gosforth
South Hylton
South Shields
St James
St Peters
Stadium of Light
Sunderland
Tyne Dock
Tynemouth
University
Walkergate
Wallsend
Wansbeck Road
West Jesmond
West Monkseaton
Whitley Bay
Location
2011/12
2012/13
Total

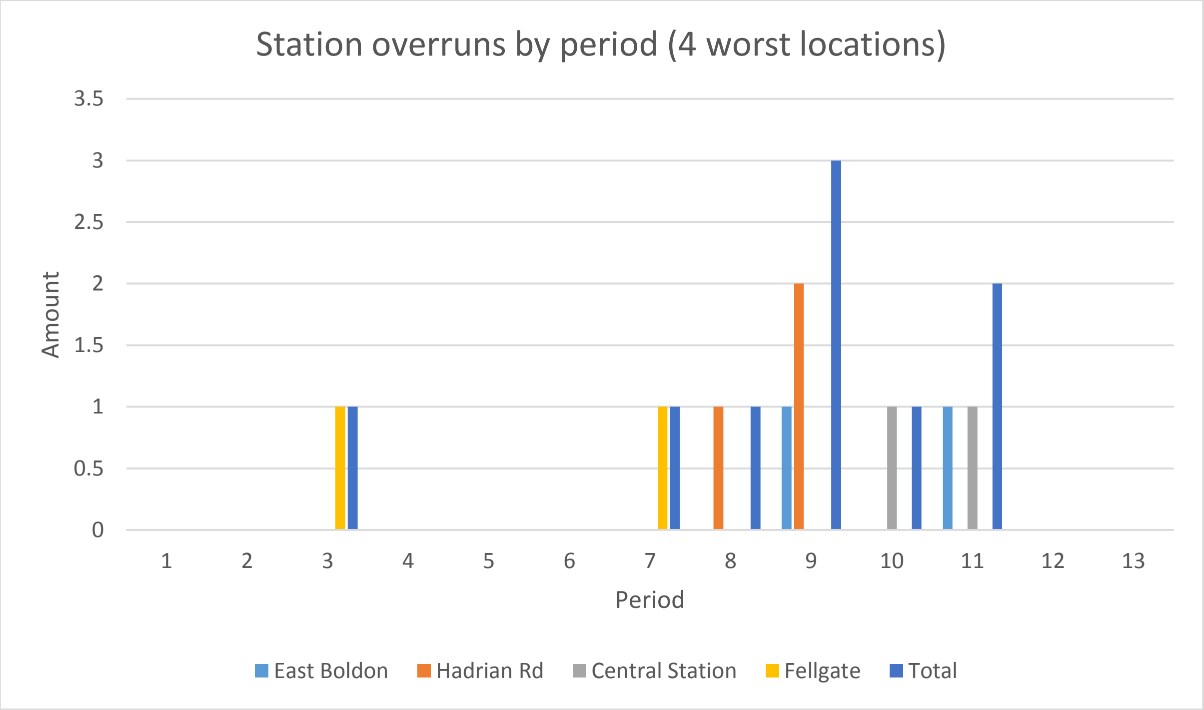
# Appendix D. Overspeeding incidents at 4 worst performing locations



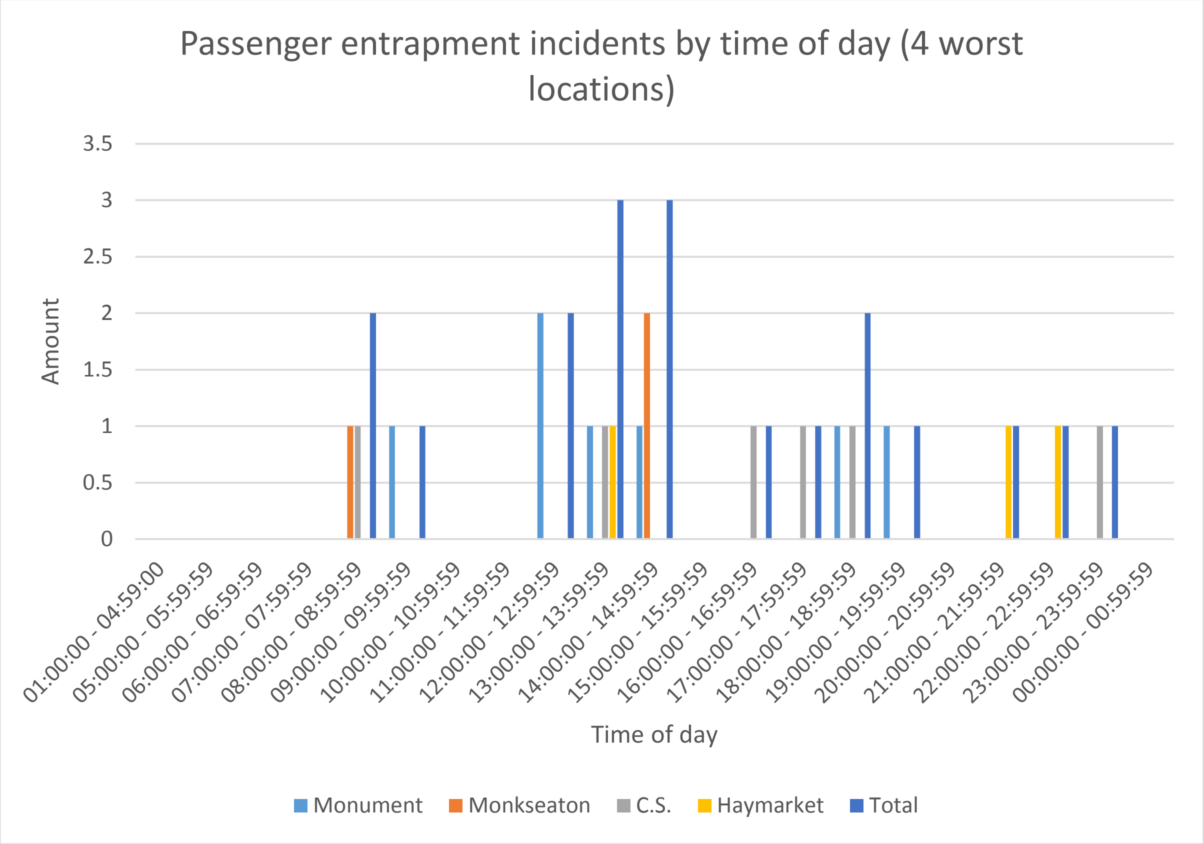


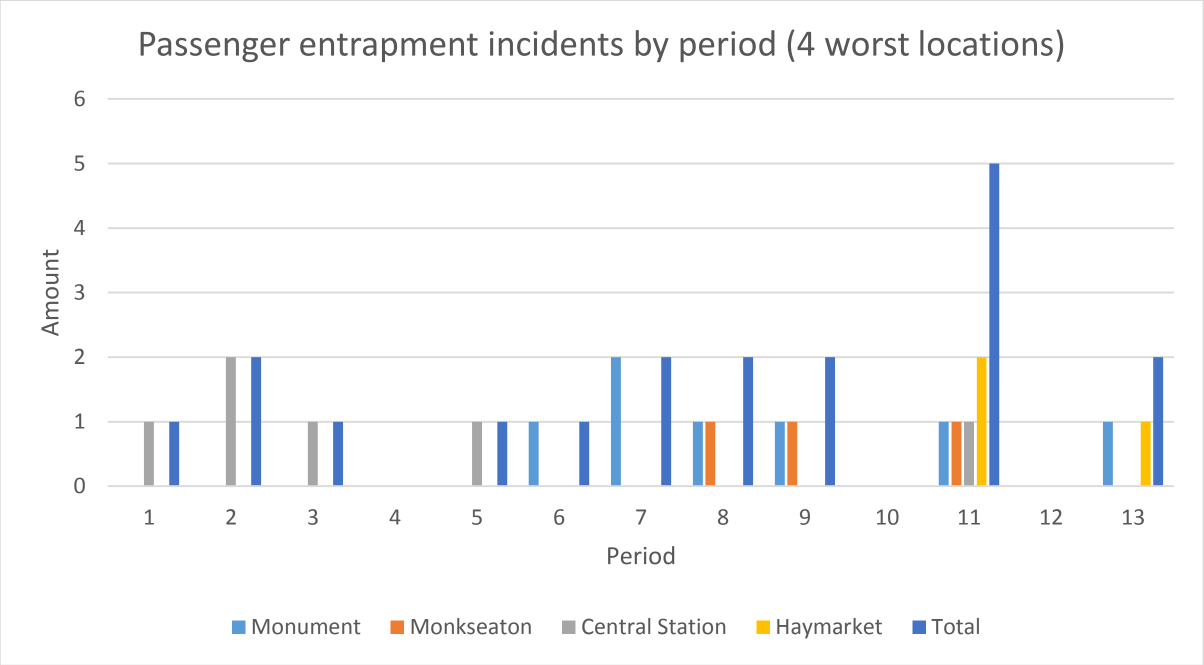
# Appendix E. Station overruns at 4 worst performing locations



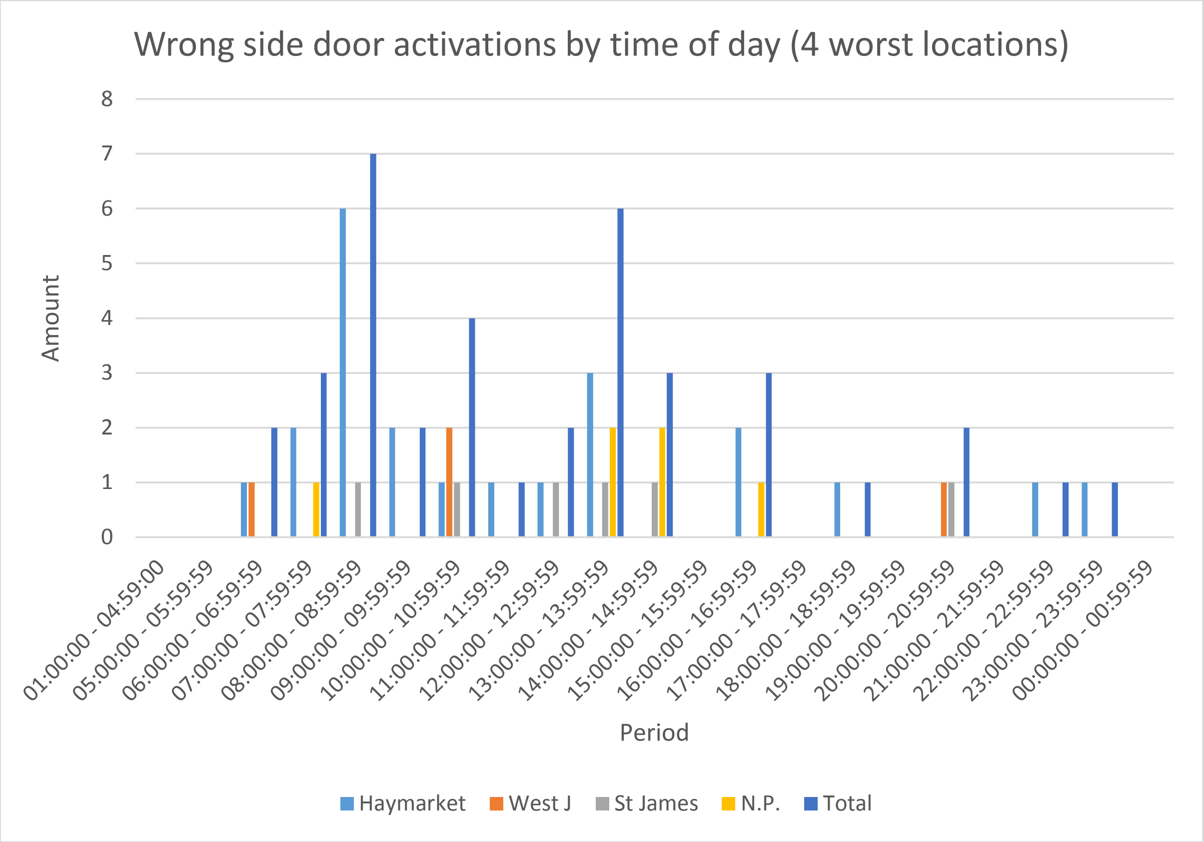


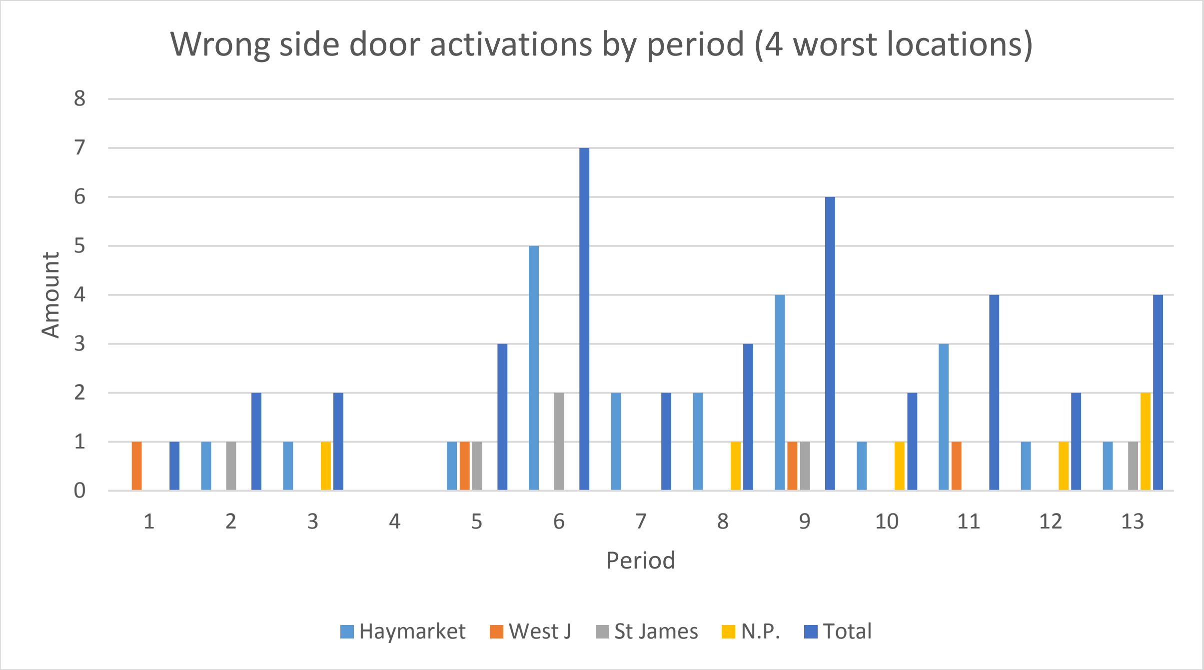
# Appendix F. Passenger entrapment incidents at 4 worst performing locations





# Appendix G. Wrong side door activations at 4 worst performing locations





# Appendix H. Associations between incident types

|  | Cat A SPAD | Other SPAD | Overspeed | Failure to call | Station overrun | Passenger entrapment | Wrong side doors | Wrong route | Doors obstruction | Signal faults | Dispatch equipment | Trainfault ATP | Trackfault ATP | Foliage | Low rail adhesion | Passenger overcarrie |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cat A SPAD |  | -.204 | .124 | .049 | .057 | -.398 | -.210 | -.056 | -.230 | -.412 | -.168 | -.299 | -.146 | .078 | .449 | -.264 |
| Other SPAD | -.204 |  | -.242 | .472 | .460 | .353 | .072 | .142 | .153 | .010 | .659* | -.001 | .416 | -.166 | .478 | .245 |
| Overspeed | .124 | -.242 |  | .112 | -.231 | .023 | -.081 | .434 | -.275 | .009 | -.072 | .251 | -.321 | .103 | -.156 | .125 |
| Failure to call | .049 | .472 | .112 |  | .401 | .202 | .112 | .300 | .182 | .090 | .566* | -.090 | .600* | .046 | .300 | .185 |
| Station overrun | .057 | .460 | -.231 | .401 |  | .469 | .273 | -.026 | .390 | .009 | .666* | -.466 | .397 | -.460 | .718* | .112 |
| Passenger entrap. | -.398 | .353 | .023 | .202 | .469 |  | .549 | .545 | .578* | -.100 | .633* | -.255 | .079 | -.522 | .229 | -.139 |
| Wrong side doors | -.210 | .072 | -.081 | .112 | .273 | .549 |  | .295 | .486 | .040 | .474 | .083 | .060 | -.387 | .090 | .240 |
| Wrong route | -.056 | .142 | .434 | .300 | -.026 | .545 | .295 |  | .228 | .107 | .579* | .084 | .028 | -.229 | .228 | -.208 |
| Doors obstr. | -.230 | .153 | -.275 | .182 | .390 | .578* | .486 | .228 |  | -.157 | .475 | -.513 | .446 | -.815** | .278 | -.470 |
| Signal faults | -.412 | .010 | .009 | .090 | .009 | -.100 | .040 | .107 | -.157 |  | .034 | .635* | -.176 | .095 | -.137 | .156 |
| Dispatch equipment | -.168 | .659* | -.072 | .566* | .666* | .633* | .474 | .579* | .475 | .034 |  | -.199 | .410 | -.509 | .592* | .125 |
| Trainfault ATP | -.299 | -.001 | .251 | -.090 | -.466 | -.255 | .083 | .084 | -.513 | .635* | -.199 |  | -.409 | .522 | -.492 | .337 |
| Trackfault ATP | -.146 | .416 | -.321 | .600* | .397 | .079 | .060 | .028 | .446 | -.176 | .410 | -.409 |  | -.150 | .326 | -.096 |
| Foliage | .078 | -.166 | .103 | .046 | -.460 | -.522 | -.387 | -.229 | -.815** | .095 | -.509 | .522 | -.150 |  | -.474 | .295 |
| Low rail adhesion | .449 | .478 | -.156 | .300 | .718** | .229 | .090 | .228 | .278 | -.137 | .592* | -.492 | .326 | -.474 |  | -.159 |
| Passenger overc. | -.264 | .245 | .125 | .185 | .112 | -.139 | .240 | -.208 | -.470 | .156 | .125 | .337 | -.096 | .295 | -.159 |  |

**Table 38. Association in the period-based sample. * correlation is significant at the 0.05 level (2-tailed); **correlation is significant at the 0.01 level (2-tailed).**

| | Cat A SPAD | Other SPAD | Overspeed | Failure to call | Station overrun | Passenger entrapment | Wrong side doors | Wrong route | Doors obstruction | Signal faults | Trainfault ATP | Trackfault ATP | Dispatch equipment | Foliage | Low rail adhesion | Passenger overcarrie |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cat A SPAD | | .035 | .431 | -.302 | .498* | .574** | .601** | .165 | .385 | .207 | .524* | .138 | -.036 | .431 | .165 | .172 |
| Other SPAD | .035 | | .179 | .347 | .111 | .232 | .350 | .204 | .190 | .024 | .057 | .095 | .271 | .412 | .105 | .190 |
| Overspeed | .431 | .179 | | .104 | .248 | .772** | .453* | -.317 | .472* | .410 | .120 | .244 | -.022 | .305 | -.200 | -.124 |
| Failure to call | -.302 | .347 | .104 | | -.163 | .074 | .303 | -.024 | .082 | .328 | -.013 | .315 | .282 | .076 | -.077 | -.214 |
| Station overrun | .498* | .111 | .248 | -.163 | | .228 | .339 | .138 | .378 | .335 | .224 | .106 | .421 | .580** | .430 | -.209 |
| Passenger entrap. | .574** | .323 | .772** | .074 | .228 | | .696** | -.226 | .606** | .326 | .261 | .109 | -.065 | .338 | -.298 | -.022 |
| Wrong side doors | .601** | .350 | .453* | .303 | .339 | .696** | | .076 | .411 | .398 | .408 | .258 | .299 | .651** | .162 | .102 |
| Wrong route | .165 | .204 | -.317 | -.024 | .138 | -.226 | .076 | | -.135 | .042 | .235 | .397 | .164 | .184 | .450* | .441* |
| Doors obstr. | .385 | .190 | .472* | .082 | .378 | .606** | .411 | -.135 | | -.093 | .343 | .222 | -.024 | .528* | -.289 | -.062 |
| Signal faults | .207 | .024 | .410 | .328 | .335 | .326 | .398 | .042 | -.093 | | -0.71 | .364 | .473* | .139 | .277 | -.356 |
| Trainfault ATP | .524* | .057 | .203 | -.013 | .224 | .261 | .408 | .235 | .343 | -.071 | | .025 | -.047 | .468* | .010 | .523* |
| Trackfault ATP | .138 | .095 | .244 | .315 | .106 | .109 | .258 | .397 | .222 | .364 | .025 | | .295 | .213 | .294 | -.061 |
| Dispatch equip. | -.036 | .271 | -.022 | .282 | .421 | -.065 | .299 | .164 | -.024 | .473* | -.047 | .295 | | .366 | .581** | -.477* |
| Foliage | .431 | .412 | .305 | .076 | .580** | .338 | .651** | .184 | .528* | .139 | .468* | .213 | .366 | | .317 | .265 |
| Low rail adhesion | .165 | .105 | -.200 | -.077 | .430 | -.298 | .162 | .450* | -.289 | .277 | .010 | .294 | .581** | .317 | | -.102 |
| Passenger overc. | .172 | .190 | -.124 | -.214 | -.209 | -.022 | .102 | .441* | -.062 | -.356 | .523* | -0.61 | -.477* | .265 | -.102 | |

**Table 39. Associations by time of day. * correlation is significant at the 0.05 level (2-tailed); **correlation is significant at the 0.01 level (2-tailed).**

| | Cat A SPAD | Other SPAD | Overspeed | Failure to call | Station overrun | Passenger entrapment | Wrong side doors | Wrong route | Doors obstruction | Signal faults | Dispatch equipment | Trainfault ATP | Trackfault ATP | Foliage | Low rail adhesion | Passenger overcarrie |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cat A SPAD | | .082 | .256* | -.126 | -.059 | -.051 | -.139 | .334** | -.087 | .269* | .181 | .301* | .260* | .141 | .082 | .526** |
| Other SPAD | .082 | | .113 | .045 | -.055 | -.159 | -.010 | .032 | -.053 | .067 | .297* | .137 | -.039 | .255* | .104 | .263* |
| Overspeed | .256* | .113 | | -.053 | .095 | -.240 | .042 | -.163 | -.036 | -.093 | -.058 | .190 | .307* | -.140 | .204 | .193 |
| Failure to call | -.126 | .045 | -.053 | | .282* | -.001 | .118 | -.120 | .081 | .131 | -.023 | -.153 | -.153 | -.210 | -.076 | -.090 |
| Station overrun | -.059 | -.055 | .095 | .282* | | .273* | -.032 | -.043 | .192 | .112 | -.158 | -.129 | .121 | -.407** | .014 | -.174 |
| Passenger entrap. | -.051 | -.159 | -.240 | -.001 | .273* | | .176 | -.089 | .280* | -.031 | -.225 | .040 | .103 | -.377** | -.321* | -.257* |
| Wrong side doors | -.139 | -.010 | .042 | .118 | -.032 | .176 | | -.121 | .122 | .034 | -.064 | -.071 | .285* | -.138 | -.150 | .080 |
| Wrong route | .334** | .032 | -.163 | -.120 | -.043 | -.089 | -.121 | | .115 | .293* | -.080 | .205 | -.094 | .398** | .261* | .475** |
| Doors obstr. | -.087 | -.053 | -.036 | .081 | .192 | .280* | .122 | .115 | | -.083 | -.431** | .140 | .129 | -.226 | -.164 | -.164 |
| Signal faults | .269* | .067 | -.093 | .131 | .112 | -.031 | .034 | .293* | -.083 | | .233 | .221 | .099 | .202 | .000 | .408** |
| Dispatch equip. | .181 | .297* | -.058 | -.023 | -.158 | -.225 | -.064 | -.080 | -.431** | .233 | | .001 | -.078 | .301* | .147 | .236 |
| Trainfault ATP | .301* | .137 | .190 | -.153 | -.129 | .040 | -.071 | .205 | .140 | .221 | .001 | | .283* | .148 | .164 | .502** |
| Trackfault ATP | .260* | -.039 | .307* | -.153 | .121 | .103 | .285* | -.094 | .129 | .099 | -.078 | .283* | | -.219 | -.086 | .226 |
| Foliage | .141 | .255* | -.140 | -.210 | -.407** | -.377** | -.138 | .398** | -.226 | .202 | .301* | .148 | -.219 | | .278* | .397** |
| Low rail adhesion | .082 | .104 | .204 | -.076 | .014 | -.321* | -.150 | .261* | -.164 | .000 | .147 | .164 | -.086 | .278* | | .349** |
| Passenger overc. | .526** | .263* | .193 | -.090 | -.174 | -.257* | .080 | .475** | -.164 | .408** | .236 | .502** | .226 | .397** | .349** | |

**Table 40. Associations by location. * correlation is significant at the 0.05 level (2-tailed); **correlation is significant at the 0.01 level (2-tailed).**

# Appendix I. Questionnaire used for the survey of Tyne & Wear Metro drivers

Dear driver,

This questionnaire forms part of research aimed at identifying how the design of a metro system and its immediate physical environment affect the safety-related performance of drivers. Research on mainline railways suggests that changing and improving certain parts of the system design can reduce the number of safety-related incidents. This research aims to explore whether those findings can be applied to urban rail systems. The questionnaire should take around 10 minutes to complete.

The study is carried out by NewRail – The Centre for Railway Research at Newcastle University and is fully funded by the Institute of Sustainability at Newcastle University. We can assure you that your answers are completely anonymous and will not be shared with the management of DB Regio T&W or Nexus. Also, you can miss any questions you are uncomfortable with, although we would like you to fill it all in if possible. Any scientific articles from the data obtained here will not allow any identification of individuals.

The research has the potential to contribute to the reduction of operational risk in the Tyne & Wear Metro, and one of its aims is to provide a set of recommendations on design improvements. Your answers will help specifying areas of the system design that require monitoring and potential intervention.

If you have any questions regarding the questionnaire please do not hesitate to contact Aleks Rjabovs at the email address or phone number below.

With many thanks for your cooperation,

Aleks Rjabovs (a.rjabovs@ncl.ac.uk 07717382070), NewRail, Newcastle University,
Professor Mark Robinson, NewRail, Newcastle University,
Roberto Palacin, NewRail, Newcastle University.

Q1. Your opinions on several issues: please tick one box on each line to show how much you agree or disagree with the statement

| Statement | Strongly agree | Agree | Just agree | Not sure | Just disagree | Disagree | Strongly disagree |
|---|---|---|---|---|---|---|---|
| The safety record of the Tyne & Wear Metro in the past 3 years has improved | | | | | | | |
| Since I joined the Metro, my safety-related performance has changed for the better | | | | | | | |
| My route knowledge of the Metro is good | | | | | | | |
| My confidence reduces while driving during possessions or engineering works | | | | | | | |
| The training provided for operations in degraded mode is adequate | | | | | | | |
| The moment I enter or leave a tunnel, I feel more alert | | | | | | | |
| Running signals between the stations are easy to interact with | | | | | | | |
| Running signals at the stations are easy to interact with | | | | | | | |
| I am less alert if the outside physical environment is monotonous | | | | | | | |
| I prefer varied outside environment, such as a mix of vegetation and buildings | | | | | | | |
| The recent change in door closing procedure from 2 to 1 button sequence is easier to operate | | | | | | | |
| A 1-button sequence might increase the occurrence of passenger entrapment | | | | | | | |
| I like mirrors as station dispatch equipment | | | | | | | |
| I like monitors as station dispatch equipment | | | | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Statement** | **Strongly agree** | **Agree** | **Just agree** | **Not sure** | **Just disagree** | **Disagree** | **Strongly disagree** |
| When coming to a scheduled stop I pay attention to a running signal at the platform end | | | | | | | |
| Running signals located far from the platform end can make selection of a stopping position difficult | | | | | | | |
| It is difficult to choose which side doors to open when station signals and the platform are on opposite sides | | | | | | | |
| Signalling at ground level can be confusing after driving a train in passenger service | | | | | | | |
| The change of platform side does not affect my ability to select correct side to open the doors | | | | | | | |
| The stations differ a lot in terms of driver visibility of passengers on a platform. | | | | | | | |
| I prefer the ¾ life refurbished cab to the original one | | | | | | | |
| If the time between two stations is more than 2.5 minutes, this improves my alertness | | | | | | | |
| I feel more alert when the time between stations is less than 1 minute | | | | | | | |
| I prefer steep change in speed limits rather than gradual change | | | | | | | |
| My familiarity with operational protocols at sidings is very good | | | | | | | |
| I have good familiarity with the layout of the depot | | | | | | | |
| I find it harder to keep within higher speed limit than lower speed limit | | | | | | | |

Q2. Please give a mark out of 10 for each of these types of signals in terms of how easy they are to interact with and interpret: 10 – for the easiest/best, and 1 – the hardest.

| **Signal type** | **Mark/10** |
|---|---|
| Running signal on Network Rail infrastructure | |
| Running signal on Tyne & Wear Metro infrastructure | |
| Repeater | |
| Advance warning signal | |
| Flashing aspects | |
| Ground position lights | |
| Junction indicators | |

Q3. Please indicate in the box below the main reason for any low marks allocated above. Please mention any particular location where this signal type is hard to interpret.

Q4. Please give a mark out of 10 for each of these potential causes for wrong side doors activation. 10 – very important, 1 – not important at all.

| **Potential cause** | **Mark/10** |
|---|---|
| Attention lapse | |
| Lack of reminders for drivers at stations | |
| Layout of the door control | |
| Distractions | |
| Inadequate training | |
| Other: (please name here) | |
| Other: (please name here) | |

Q5. In the box below please name up to 3 stations that you believe have the biggest chance of wrong side door activation. Please state any reasons for selecting these stations.

Q6. Please give a mark out of 10 for each of these potential causal factors for passenger entrapment incidents. 10 – very important, 1 – not important at all.

| **Potential causal factor** | **Mark/10** |
|---|---|
| Night time | |
| Snow | |
| Rain | |
| Mist | |
| Direct sunlight | |
| Vegetation overgrowth | |
| Location of station infrastructure, e.g. CCTV cameras | |
| Overcrowding at a platform or in a train | |
| Winter clothing on passengers | |
| Shopping bags and suitcases | |
| Layout of the stations | |
| Design of passenger approaches | |
| Mobility aid equipment, e.g. walking sticks, crutches | |
| Station dispatch instructions/procedures used in the Metro | |
| Low height passengers, e.g. kids | |
| Other: (please name here) | |
| Other: (please name here) | |

Q7. In the box below please name up to 3 stations that you believe have the biggest chance for a passenger entrapment incident to occur. Please give reasons for selecting these stations.

Q8. Thinking about the "3/4 life" cab; please state how satisfied you are with its components.

| **Component** | **Strongly satisfied** | **Satisfied** | **Just satisfied** | **Not sure** | **Just dissatisfied** | **Dissatisfied** | **Strongly dissatisfied** |
|---|---|---|---|---|---|---|---|
| Driver's seat | | | | | | | |
| Driver's desk | | | | | | | |
| FASSI | | | | | | | |
| Driver's Safety Device (DSD) | | | | | | | |
| Deadman's Vigilance Device (DVD) | | | | | | | |
| Master controller | | | | | | | |
| Heating, ventilation and air conditioning unit | | | | | | | |
| Other: | | | | | | | |

Q9. In your opinion, what season has the highest amount of passenger entrapment incidents? Please select one. *Please tick one appropriate answer.*

☐ Winter ☐ Spring ☐ Summer ☐ Autumn ☐ Don't know

Q10. Please assess the following statements and tick one appropriate answer in each row.

| | **From the start of the fixed driving portion, in minutes** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Statement** | **Under** | **31-60** | **61-90** | **91-120** | **121-150** | **151-180** | **Over 181** | **Never** |
| My alertness start to decrease | | | | | | | | |
| My boredom starts | | | | | | | | |
| Most incidents are likely to happen | | | | | | | | |
| Least incidents are likely to happen | | | | | | | | |

Q11. In the past 3 years have you ever been involved in any of these types of incidents? *Please tick all that apply*

☐ Category A SPaD
☐ Category B/C/D SPaD
☐ Overspeeding
☐ Station overrun
☐ Failure to call
☐ Passenger entrapment
☐ Wrong side doors activation

Finally, three questions about you personally:

Q12. What is your gender? *Please tick one appropriate answer.*
☐ Male ☐ Female

Q13. Please tick your age group
☐ 18-25 ☐ 26-35 ☐ 36-45 ☐ 46-55 ☐ 55+

Q14. How long have you been a metro driver? *Please tick one appropriate answer.*
☐ Less than 3 years ☐ 3 to 10 years ☐ 10 to 20 years ☐ More than 20 years

Q15. Please use the box below to write any comments you would like to raise about any of the questions or issues in the survey; your opinions are appreciated.

With many thanks for filling this in. We will provide a summary of the findings for you all when the research is completed.

Aleks Rjabovs, Professor Mark Robinson and Roberto Palacin

# Appendix J. Screenshots of virtual Tyne & Wear Metro (scenario 1)


mile/h


mile/h


mile/h

mile/h

Heworth
mile/h


50


50


Gateshead Stadium


d Stadium


Pelaw


Heworth
S


50
Felling

# Appendix K. Screenshots of virtual Tyne & Wear Metro (scenario 3)


Pelaw


great north forest
great north forest


mile/h
0
10
20
30
40
50
60
70
80
90
100
110
120


TICKETS
mile/h
0
10
20
30
40
50
60
70
80
90
100
110
120


mile/h


Gateshead Stadium
mile/h


Gateshead Stadium
mile/h


Gateshead Stadium
mile/h

# Appendix L. Simulation assessment questionnaire

1. Please rate the following aspects of the simulation

| | Very good | Good | Don't know | Poor | Very poor |
|---|---|---|---|---|---|
| Visuals | | | | | |
| Train handling | | | | | |
| Immersion | | | | | |
| Comfort | | | | | |

2. Was it easy to notice changes introduced to the physical environment?
   ☐ Yes ☐ No ☐ Can't tell

3. Would you consider this simulation as a good representation of driver's duties in T&W Metro?
   ☐ Yes ☐ No ☐ Can't tell

4. What can be improved in order to make it more realistic?