Journal of Rail Transport Planning & Management 23 (2022) 100332

Contents lists available at ScienceDirect

Journal of Rail Transport Planning & Management

journal homepage: www.elsevier.com/locate/jrtpm

What employees do today because of their experience yesterday: Previous exposure to yellow:number aspects as a cause for SPAD incidents

Julia Burggraaf[a,b,*], Jop Groeneweg[a,c,d], Simone Sillem[a], Pieter van Gelder[a]

[a] *Safety and Security Science Group, Delft University of Technology, Jaffalaan 5, 2628 BX, Delft, the Netherlands*
[b] *ProRail, Moreelsepark 3, 3511 EP, Utrecht, the Netherlands*
[c] *Cognitive Psychology Unit, Leiden University, Wassenaarseweg 52, 2333 AK, Leiden, the Netherlands*
[d] *TNO Leiden, Schipholweg 77-89, 2316 ZL, Leiden, the Netherlands*

ARTICLE INFO

Keywords:
SPAD
Incidental learning
Train driver
Signal
Rail safety
Human factors

ABSTRACT

When a train passes a red aspect, this is called a Signal Passed at Danger event or SPAD. Sometimes it is easy to identify the SPAD cause but in other cases it is unclear why the incident occurred, especially if the system operated as usual and the train driver was trained and experienced just like his or her colleagues. In previous research, train driver deceleration behaviour has been shown to be influenced by frequent exposure in the previous 14 days to less restrictive and visually similar signal aspects in the same location. Previous exposure can contribute to SPAD causation unless the initial insufficient deceleration is corrected in time. Six years of SPAD data and red aspect approaches in the Netherlands was used to test whether previous exposure to yellow:number aspects corresponds with a statistically significant increase in SPAD incidents if there is a small window for correction available to drivers. The permitted track speed and signal distance influence the size of this window. The results provide evidence for previous exposure as a cause for SPADs and details to identify locations with increased SPAD probability. Changes in infrastructure and timetable design or adding safety measures for these locations can prevent future SPADs.

1. Introduction

When a train passes a red aspect without authorisation, this is called a Signal Passed At Danger event, or SPAD. SPADs receive a lot of attention because they can lead to severe consequences, for instance a train collision. Additionally, even non-harmful SPADs in terms of injury and damage still have direct and indirect costs (Kyriakidis et al., 2019). New technical systems like the European Rail Traffic Management System (ERTMS) have been implemented and are expected to provide additional SPAD prevention (Ministry of Infrastructure and the, 2013). However, in the Netherlands, the implementation of ERTMS on a national level may take up to 30 years and safe implementation is complex (Smith et al., 2012; Ministerie van Infrastructuur en Waterstaat, 2020). At the same time, Dutch rail is predicted to become busier and busier. The Infrastructure Manager aims to support a growth in passenger transport of 30% by 2030 and 45% in freight transport by 2030 (BNR Webredactie CEO ProRail, 2020; GWW Grotere capaciteit spoorgoederenvervoer in,

* Corresponding author. Safety and Security Science Group, Delft University of Technology, Jaffalaan 5, 2628 BX Delft, the Netherlands.
E-mail address: j.m.burggraaf@tudelft.nl (J. Burggraaf).

https://doi.org/10.1016/j.jrtpm.2022.100332
Received 26 October 2021; Received in revised form 4 February 2022; Accepted 13 June 2022
Available online 24 June 2022
2210-9706/

2030). This increase requires changes to the infrastructure and the timetable. In order to keep on improving the level of safety, prevent SPADs and improve the system as a whole, it is important to understand which factors increase or decrease the probability that a SPAD occurs.

When SPADs occur, sometimes the cause is easy to identify, but on other occasions it remains unclear what the exact cause was. This is especially the case when SPADs do not occur because of a technical failure but because a mistake was made in driving behaviour. Whilst 'insufficient deceleration' is often identified as a cause of a SPAD, it is harder to identify why the train driver did not decelerate sufficiently (Gibson et al., 2021). Different experts have also been shown to regard different factors as having the greatest influence on the occurrence of the same incident, based on their interpretation of the same incident report (Baysari et al., 2008; Madigan et al., 2016).

The field of Human Factors looks into the factors that influence human behaviour and can increase or decrease the probability of an error (Wickens et al., 2003). Human factors research in rail has become more common since the mid to late 1990s (Wilson and Norris, 2005). Visibility of signals is one factor that is often considered within rail. Another example of a factor is signal placement, with a focus on whether it might be confusing for the train driver to know which signal is for him/her (Wilson and Norris, 2005; Hamilton and Clarke, 2005; Lawton and Ward, 2005; Naweed et al., 2019; Stanton and Walker, 2011). Route knowledge can help to prevent mistakes as a result of problems in signal visibility or interpretation (Punzet et al., 2018; Naweed et al., 2013). Another option is to improve placement and make the infrastructure more logical from the train driver's perspective. Improving the placement of signals is a more fundamental solution in contrast to route knowledge, where there is more variability with respect to its implementation (e.g. when the driver last drove in that location and the level of knowledge that was gained and will be sustained under stressful situations). In essence, infrastructure changes are a more reliable safety barrier than interventions related to training and awareness. It therefore makes sense to consider whether the infrastructure can be adjusted to decrease the probability of human error.

In rail, the infrastructure (including the signalling system) is an important part of the working environment of the train driver. We thus advocate improving that working environment. Often, the effect of the working environment of the train driver at the moment of his/her task execution is considered. This is a core question within human factors: 'which factors (at moment x) influence the performance at moment x?' And often in incident analysis: 'which factors (at moment x) caused the error at moment x?' But what about the past? Does it matter what an employee experienced yesterday or last week or last month? The role of experience is often considered in incident analysis in such a way that a lack of experience is considered as a potential error-promoting factor. It is however not considered that previous experience or previous exposure can also be an error-promoting factor in certain situations (Wiegmann and Shappell, 2003; Gertman et al., 2004). The question that is not asked: 'do factors (at moment x-Δt) cause an error at moment x?'

In this research we investigate the possibility that learning through exposure increases the probability of an error. Learning is the creation of new pathways in our brain and also the strengthening of existing pathways (Gazzaniga et al., 2009). People learn every day, both intentionally and incidentally. For skilled work, a lot of learning takes place on the job. This learning does not stop once we are able to perform a skill. Consider the example of getting a new job and driving the new route to work. The next time we drive the same route, the neural pathways for driving that route are strengthened. With time, we need very little attention to turn in the correct places. A 'driving to work' schema has developed in our brain.

Within psychology, schemas such as the 'driving to work' schema refer to unconscious mental structures that guide cognitive processes and behaviour. The development of schemas for activities is very useful as these schemas allow us to perform the activities with less attention (Bartlett, 1932; Piaget and Cook, 1952; Neumann et al., 1984). The schemas can be activated by conscious intention to perform an action, but also by (visual) cues. This latter activation is called bottom-up activation. A cue can be visual, but also auditory, tactile or a previous action. However, problems can occur if a schema is activated and performed in a situation where that behaviour is not suitable. This can happen if the contextual cues are similar to those situations where one usually performs the behaviour (Norman, 1981). In line with this thinking, it follows that one's behaviour in a current situation can be influenced by one's past behaviour in (visually) similar situations. The environment one is exposed to, both in terms of cues and required behaviour, determines whether the learning increases the probability of an error.

Balfe and Doyle analysed a multi-SPAD signal in Ireland using the RSSB SPAD Hazard Checklist. The SPADs were preceded by a yellow aspect at a signal which usually showed a double-yellow aspect. The double-yellow aspect which shows in the standard situation indicates a red aspect at the end of the nearest loop platform, two signals further, whilst the yellow signal (present during the SPADs) indicates that the next signal has a red aspect. During all three SPAD events at the signal in front of the platform, the drivers reported that they understood the previous signal to be showing double yellow while it had in fact shown a single yellow aspect. Balfe and Doyle therefore identify the fact that the aspect is frequently double-yellow in that specific location as a potentially contributing factor to the SPADs (Balfe and Doyle, 2021). The hypothesis is that a specific aspect (yellow in the above example) activates the behaviour belonging to a different aspect (double-yellow in the above example). If the signal that showed the single yellow aspect during the SPAD event would have shown a green aspect on most previous approaches instead of the double-yellow aspect that was shown frequently, the SPADs might not have occurred.

In Dutch rail, double-yellow aspects are not used, but other yellow aspect variations are common. Burggraaf and colleagues identified specific types of red aspect approaches during which the train driver's behaviour at a given day is indeed affected by the type of yellow aspect that was present in that location over the previous 14 days (Burggraaf et al., 2021). It is important to understand whether those measured changes in behaviour can cause SPADs, in which situations and to what extent.

The section below explains during which red aspect approaches in Dutch rail the wrong schema might be activated and why this initial error may or may not lead to a SPAD. The train driver's 'window for correction' is introduced, followed by the methods section, results, conclusion and discussion.

1.1. Situations in Dutch rail with potential of wrong schema activation

In Dutch rail, a red aspect is most often part of a green-yellow-red aspect sequence. There are, however, also yellow aspect variations apart from the 'pure' yellow aspect. The yellow aspect can be accompanied by a number (e.g. 4, 6, 8, 13) that indicates the speed limit present at the next signal (respectively 40, 60, 80 and 130 km/h). The signal in Fig. 1 indicates that the train driver should decelerate to 40 km/h and pass the next signal with a speed of 40 km/h or less.

These yellow aspect variations can be part of a red aspect approach (e.g. green – yellow:4 – yellow – red). They can also be used to reduce the speed before a switch. A specific signal can often show the same yellow:number aspect if it is part of a speed reduction before a switch or if it is part of a red aspect approach towards a station platform. Fig. 2 shows a situation where the signal at the station stop has a red aspect and is preceded by a yellow and a yellow:8 aspect. These locations are of specific interest because red aspect approaches occur more often at the same signals when the red aspect is at a signal at a station stop, than when the signal is situated along regular tracks.

Fig. 3 shows a location where the first signal can have the aspect yellow:8 because of a switch (a in Fig. 3), because of a red aspect at the scheduled stop for this train (b in Fig. 3), or because of a red aspect that is not at a scheduled stop (c in Fig. 3). The same signal can also occasionally show the aspect yellow instead of yellow:8 when the next signal has a red aspect (d in Fig. 3).

During the yellow-red approach, the deceleration behaviour suitable for the yellow:8 approach (see a, b and c in Fig. 3) might be activated despite having passed a yellow aspect (see d in Fig. 3). This will lead to insufficient deceleration. Although the aspects yellow and yellow:8 are very different in meaning, they have visual similarities which can be enough to trigger the behaviour normally followed by yellow:8, if that is a common aspect in that specific location (Burggraaf et al., 2021). Visual similarity between two objects is defined by the amount of shared points or common features and the type of difference. When the difference is a deletion at end points (such as the letter 8 not showing at the bottom of the signal), then the visual similarity is higher than when there are deletions that lead to breaks in continuity or when there are mirror image reversals (Singer and Lappin, 1976).

In the above example yellow:8 is used, but the effect is hypothesised to be present for any yellow:number aspect. The probability of the train driver displaying the past behaviour increases if the train driver has passed that signal with the yellow:number aspect more often.

1.2. When wrong schema activation leads to a SPAD

In order for a SPAD to occur because of wrong schema activation in the brain, four elements should be present:

1. The train driver performs the deceleration behaviour suitable in previous situations
2. There is a non-negligible difference in the current required behaviour and the previous required behaviour
3. The train driver does not correct his or her behaviour in time
4. Technical (warning) systems do not intervene to correct the behaviour

1.2.1. The train driver performs the past behaviour

The previous behaviour will be performed if it is activated sufficiently due to visual similarity and frequent past exposure (bottom-up) and the behaviour is not prevented top-down by our 'will' via our supervisory attentional system (Norman and Shallice, 1986). Burggraaf and colleagues found that during the red aspect approach as part of the green-yellow-red sequence (last approach in Fig. 3), the train driver behaviour was indeed affected if the same train series had often passed a yellow:number aspect in the same location in the previous 14 days (Burggraaf et al., 2021). A higher frequency in the previous 14 days led to an increase in the change of behaviour. This effect decreased after the frequency of a yellow:number aspect in the previous 14 days exceeded 400 times. This research indicated that train driver behaviour is indeed affected by previous yellow:number aspects during yellow-red approaches.

1.2.2. There is a non-negligible difference in the required behaviour and past behaviour

Wrong schema activation will not cause a SPAD if a slight deceleration is sufficient during both previous and present approaches (for example, a deceleration of 0.31 m/s^2 is required in the previous scenario and 0.32 m/s^2 in the current approach). This minimal

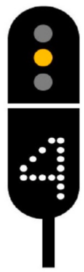

Fig. 1. Signal with yellow:4 aspect.

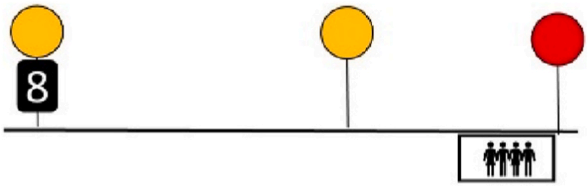

Fig. 2. Red aspect at station platform preceded by yellow and yellow:8 aspect.

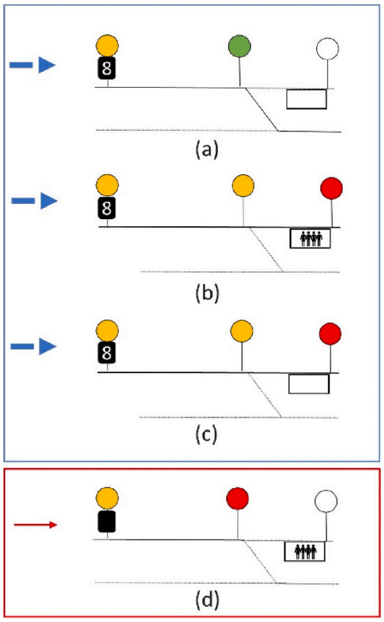

Fig. 3. Frequent yellow:8 aspects due to a switch (a), red aspect at scheduled stop (b) or red aspect without scheduled stop (c). Increased probability of error in bottom situation where the aspect is yellow instead of yellow:8 in the same location, followed by a red aspect (d).

difference can occur when the track speed is low and/or the distance between the signals is large.

1.2.3. The train driver does not correct his/her behaviour in time

A SPAD can be prevented if the train driver corrects his or her behaviour in time. The theory of wrong schema activation predicts an initial insufficient deceleration after passing the yellow aspect. The train driver can realise the mistake and start to decelerate forcefully upon seeing the red aspect. Factors that increase or decrease the probability that a driver will be able to correct his or her behaviour in time are expanded upon in section 1.3.

1.2.4. Technical (warning) systems do not intervene to correct the behaviour

A SPAD can also be prevented if other preventative interventions are present. The Dutch Railways has implemented an auditory warning system called Orbit which is designed to prevent SPADs by warning the train drivers with an auditory message when they approach a red aspect at a higher speed than desired. This warning system is not yet installed on all trains, nor operational for all signals. ERTMS can also provide preventative intervention but is not nationally implemented yet in the Netherlands (Ministry of Infrastructure and the, 2013).

1.3. Investigating element three: the train driver does not correct his/her behaviour in time

The four elements that need to be present for wrong schema activation to cause a SPAD have been described above. Previous research has shown an effect of past yellow aspect exposure and thus the presence of element one. The risk of a SPAD is limited to those locations where there is a difference in past and current required behaviour (element two), otherwise the past behaviour is not erroneous, and where there are no technical (warning) systems that can intervene (element four). The remaining question is whether train drivers are able to correct their own behaviour in time (element three) and prevent the wrong schema activation from actually causing a SPAD.

Whether the train driver will still be able to stop in front of the red aspect depends on the moment at which the train driver sees the red aspect, the deceleration power of the train and the size of the window for correction. The top of Fig. 4 shows a situation where the track speed is 130 km/h and the aspect at the first signal is often yellow:8, indicating a speed reduction to 80 km/h. The distance between the first and second signal is 1095 m. In order for the train to drive at 80 km/h at the next signal, as is required, the train must decelerate continuously at 0.37 m/s^2. At the bottom of Fig. 4 we have the hypothetical situation where the train passes a yellow aspect and decelerates at 0.37 m/s^2, as is commonly suitable in this location. However, since the aspect is yellow and not yellow:8, a continuous deceleration of 0.37 m/s^2 is insufficient to stop in front of the red aspect. In this hypothetical example, the train driver sees the red aspect at 300 m. He or she then attempts to correct his or her initial insufficient deceleration. If the train is able to continuously decelerate by at least 1.19 m/s^2 from that point onwards, then a SPAD can be prevented. Given that the emergency brake power of most passenger trains is around 1.2 m/s^2 and the train driver needs to initiate the emergency brake after perceiving the red aspect, this situation is one where there is a very small window for correction, assuming correct perception of the red aspect at 300 m.

It is unknown how often a train driver perceives a red aspect at 300 m. The Dutch rail infrastructure manager ProRail has directives on the minimal distance at which a signal needs to be visible. At track speeds below 80 km/h, the minimal visibility distance is 200 m. At 100, 130 and 140 km/h the minimal visibility distances are respectively 250, 325 and 350 m to maintain visibility for 9 s per speed (Ontwerpvoorschrift, 2020). Even if the signal is theoretically visible, this does not necessarily mean that the train driver sees the red aspect and perceives it correctly. Seeing and perceiving the red aspect in time can also be influenced by a train driver's previous experience. Summerfield and Egner state in their review on visual cognition that visual detection and recognition are guided by one's prior knowledge of what is likely to occur (Summerfield and Egner, 2009). Other factors can also influence early detection and recognition of the red aspect, like the visual conspicuity of the signal and aspect, weather conditions and situations where it is not immediately clear which signal belongs to one's track.

1.4. Research question and hypotheses

To further increase our knowledge of causes of SPADs, we investigate the following research question:

Does the frequency of a yellow:number aspect in the previous 14 days increase the probability of a SPAD if the window for correction is small?

This research question is answered for yellow-red aspect approaches where: the yellow aspect is in the same location as the yellow: number aspect, there is a difference between the previous required and current required deceleration behaviour, and there are no technical systems present to correct the behaviour.

We hypothesise that train drivers are able to correct their behaviour and prevent a SPAD, if the infrastructure provides a large window for correction. However, when the window for correction is small, we hypothesise that wrong schema activation can contribute to SPAD causation.

Even though it is unknown when the train driver perceives the red aspect, it is possible to compare signal locations on the size of window for correction by comparing how much deceleration is necessary at a specific distance in front of the signal. Categorisation as 'small', 'medium' and 'large' windows for correction is discussed in section 2.4. This research will therefore also give more insight into which sizes of window for correction are large enough to provide the opportunity for self-correction.

2. Methods and materials

To answer the research question, we investigated whether there were more SPADs by passenger trains in the Netherlands between 2014-01-01 and 2019-12-31 than can be expected based on the number of red aspect approaches. Six years of data were used to have as much data as possible in manageable quantities. Data from 2020 was not used because of a large difference in the number of red aspect

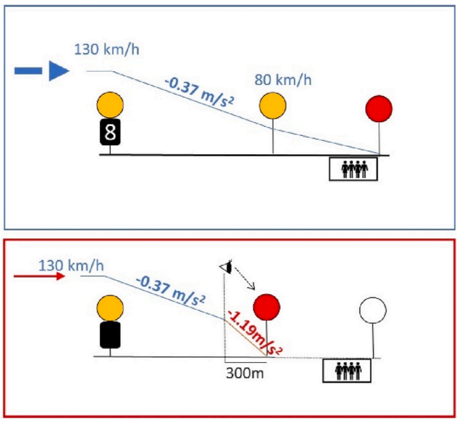

Fig. 4. Hypothetical example where the standard situation with a deceleration of 0.37 m/s^2 is sufficient. If the train driver employs the same amount of deceleration during the green-yellow-red approach, then this insufficient deceleration can be corrected if, upon seeing the red aspect at 300 m, the train can decelerate by at least 1.19 m/s^2 continuously.

approaches due to timetable adjustments during the COVID-19 pandemic. For each SPAD and red aspect approach falling within the inclusion criteria, the frequency was calculated of a yellow:number aspect in the previous 14 days in the same location and the size of the window for correction was calculated.

2.1. Data

Two main data sources were used: 1. SPAD incidents, and 2. Red aspect approaches. A list of SPAD incidents was provided by ProRail, the Dutch Rail infrastructure manager. Data of red aspect approaches was also provided by ProRail. When a train passes a signal, and the next signal is red at that point in time, it is recorded as a red aspect approach. The point in time at which a signal turns red or not-red is recorded for many of the signals on the Dutch rail network. For some signals, this point in time is not recorded, but can be deduced based on the moment that a train enters sections between signals. For some signals, this data is absent and therefore both SPADs and red aspect approaches at these locations were not included.

Whilst train kilometres are easier to obtain and therefore historically used more often, red aspect approaches are a better measure for the opportunity of SPAD occurrence (Gibson et al., 2017; Harrison et al., 2021). The red aspect approach data also provides the additional details needed to test the hypotheses.

2.2. Types of SPADs

SPADs can occur for a multitude of reasons. This research focuses on driver error as the immediatecause of SPADs and not on technical causes. The SPAD incident list was therefore filtered on SPADs that did not have a technical cause, or where the signal was put on red by the train traffic controller after the train had already passed the preceding signal. These SPADs could be excluded by only selecting the SPADs categorised as 'non-technical – other'.

The SPAD list did not include sufficient information in easily accessible format and was therefore enriched with data from the Red Aspect Approaches (RAA) dataset. The data was automatically matched based on date, train number and signal number. Only 47% of the selected SPADs could be matched with the RAA dataset. Upon inspection of the cases that were not matched, there were valid reasons why the match could not occur:

- The SPAD did not occur at a signal but at a sign
- The SPAD occurred when the train left the station whilst the departure signal was red, thus being a departure through red aspect and not a red aspect approach
- The SPAD did not occur with a passenger train but a road rail crane
- The SPAD occurred during shunting

The above situations are all outside the scope of this investigation. There were twelve SPADs that could not be matched because they were at a signal of which the red aspect time is not automatically recorded and could not be deduced. The twelve SPADs were at six different signals with six SPADs having occurred at one signal. These SPADs were within the scope, but not included in the analysis because they could not be matched. Including these SPADs manually was not an option, since the accompanying red aspect approaches should then also be added, which was not feasible.

2.3. Inclusion criteria for SPADs and red aspect approaches

The SPADs and red aspect approaches were included if they fit the criteria below.

Criteria to only select SPADs and red aspect approaches that are part of the hypothesis:

- The red aspect was part of a green-yellow-red aspect sequence
- The train was expected to approach with a speed higher than 80 km/h. The effect of exposure to past yellow:number aspects was only tested at speeds above 80 km/h because at low speeds the difference between past and required deceleration behaviour tends to be small. Speed above 80 km/h is filtered in via 1) The track speed was higher than 80 km/h according to permanent traffic signs. 2) The train did not pass a yellow signal before the red signal approach as part of a previous red aspect approach. Previous yellow aspects would have already resulted in lower train speed. 3) The train was driving before passing the yellow signal instead of departing from a station.

Criteria to avoid other effects influencing the analysis:

- The red signal was not at a scheduled stop location. These approaches were excluded because the train driver would need to stop at these locations regardless of the aspect colour.
- The above criteria related to speed excluded situations where two platforms were situated directly behind each other, thereby excluding situations where the red aspect was at the first platform whilst the train driver usually stops at the second platform.
- The yellow signal was near a station stop as defined by being part of a red aspect approach to a scheduled stop at least once.

- The auditive warning system Orbit was not present or operating on the train. This warning system helps to prevent SPADs. Since not all trains or all signals are protected by Orbit yet, it is still relevant to understand whether previous exposure to yellow:number aspects can contribute to SPADs when this warning is system is not present or operating correctly.
- For the statistical analysis which did not include the window for correction, only SPADs and red aspect approaches with a mean deceleration above 0.5 m/s^2 were included. The mean deceleration (m/s^2) was calculated via 0.5 × track speed^2 (m/s)^2/distance between signals (m). The train driver behaviour analysis by Burggraaf and colleagues showed a difference in driver behaviour for approaches where the mean deceleration was higher than 0.5 m/s^2 (Burggraaf et al., 2021). For the statistical analysis which included the window for correction, the filter on mean deceleration is replaced by categorisation based on the size of the window for correction.

2.4. Independent variables

The two independent variables were: A. the frequency of yellow:number aspect in the last 14 days for this train series and B. the window for correction. The frequency was calculated by counting the number of times the same train series passed the yellow:number signal in the last 14 days. Trains have the same train series when they are scheduled to drive the same route with the same stops. Approaches are counted if (per train series) all have the same yellow:number aspect or they do vary in yellow:number aspects with different numbers but the highest aspect frequency in the previous 14 days is above 100 and the other aspect frequencies are below 5.

The window for correction was calculated by taking into account what the yellow:number aspect was in the previous 14 days, what the distance was between the signals, and the permitted track speed. Since we do not know exactly what deceleration behaviour the train driver usually employs, we calculate what the sufficient continuous deceleration is during the yellow:number approach. This is calculated via $\frac{v^2_{track\ speed} - v^2_{aspect\ speed\ limit}}{2 * distance\ between\ signals}$. We then calculate how fast the train should decelerate at 300 m from the red aspect, if the train has been continuously decelerating with this deceleration level up until that point, via $\frac{v^2_{track\ speed} - sufficient\ continuous\ deceleration * 2 * (distance\ between\ signals - 300)}{2 * 300}$. We call this value ‘required deceleration upon 300 m’. In the example that was depicted in Fig. 4 with a track speed of 130 km/h (36.1 m/s), aspect speed of 80 km/h (22.2 m/s) and a signal distance of 1095 m, the sufficient continuous deceleration is 0.37 m/s^2 and the required deceleration upon 300 m is 1.19 m/s^2.

We categorised a required-deceleration-upon-300 m value of less than 0.6 m/s^2 as a large window for correction, since this deceleration value is easily attained and very common. A value between 0.6 and 0.8 m/s^2 is categorised as a medium window for correction. A value above 0.8 m/s^2 is categorised as a small window for correction. This categorisation is relative rather than absolute. It distinguishes locations with a larger window for correction from those with a smaller window for correction, rather than defining ‘large’ and ‘small’.

2.5. Analyses overview

Two analyses were run. The first one was performed to test whether the results of Burggraaf and colleagues could be replicated using incident data instead of behavioural data (Burggraaf et al., 2021). In this test, the window of correction was not included and the criterion for mean deceleration was included as described in section 2.3. The sample included 29 SPADs and 1,139,665 red aspect approaches.

The second analysis included the window for correction measure, which divided the SPADs and red aspect approaches into large, medium and small windows for correction. The window for correction could not be calculated for those approaches where there was no yellow:number aspect in the previous 14 days. These approaches were therefore not included in this test. A test was performed per window for correction to test the hypothesis that the relationship between frequency and chance of a SPAD only exists when there is a small window for correction. The samples included 0 SPADs and 777,510 red aspect approaches for the large window for correction, 3 SPADs and 319,533 red aspect approaches for the medium window and 17 SPADs and 54,462 red aspect approaches for the small window.

2.6. Statistical analysis

To test the effect of the yellow:number aspect frequency in the previous 14 days on the SPAD frequency, the same statistical method was used as Burggraaf and colleagues used to be able to see the exact shape of the relation between frequency and SPAD occurrence (Burggraaf et al., 2021). The bin sizes were based on that previous research, which used driver behaviour instead of SPADs. Due to technical reasons and to the large amount of data, it was necessary to select bins beforehand. It was therefore not possible to leave the frequency as an interval variable, which could potentially have been tested with a logistic regression. This was not considered a major issue since other testing is available and the previous study using driver behaviour instead of SPADs showed assumption violations to perform the logistic regression anyway. Since the relation between the yellow:number aspect frequency and train driver behaviour showed a slightly inverted u-shape in that previous study and there were many approaches in the bins of lower frequencies, the following frequency bins were chosen: 0 times a yellow:number aspect in the last 14 days (only for the first analysis), 1–50 times, 51–150 times, 151–250 times, 251–350 times, 351–450 times, 451–550 times, >550 times.

The *p*-value was calculated per bin by comparing the observed number of SPADs with the number of SPADs that is expected for the bin under the H0 assumption that there was no difference between bins. The analyses were run in R, version 3.6.2. No additional packages were used for the analyses. The R Code is provided in Appendix A. The steps are clarified with an example in Fig. 5.

In the Results section the exact *p*-values were recorded when they were below 0.05, and were listed as $p < 0.001$ when they were below 0.001. *p*-values above 0.05 were recorded as non-significant (N.S.).

Since there are relatively few SPADs, it is possible to have zero SPADs in a given bin. If that bin is not significant, it is possible that: A. there is no difference in SPAD probability, B. the probability in that bin is lower but the result is not significant due to low power, or C. the probability in that bin is in fact higher but due to a low number of red aspect approaches in the given bin, no SPADs occurred. If the average proportion, for example, is $1*10^{-5}$ SPAD per red aspect approach, then a bin with zero SPADs per 200,000 red aspect approaches is more likely to be non-significant due to option A or B since 1/200,000 is $0.5*10^{-5}$. A non-significant bin with zero SPADs per 30,000 red aspect approaches is more likely to be non-significant due to option C since 1/30,000 is $33*10^{-5}$. Non-significant bins with zero SPADs should therefore not be interpreted as low-risk. Multiple two proportion z-tests or Fisher Exact tests were not used because the desired comparison was not to test whether two bins deviated, but whether a bin could be from the overall average, including that bin, violating the assumption of independence.

A chi-square test for independence was considered as an overall test before the described test, but was not possible because of the violation of the assumption that the expected value per cell should be 5 or more in at least 80% of the cells, and no cell should have an expected value of less than one. The Fisher Exact test provided an out-of-workspace error in R due to the large number of variables and number of red aspect approaches. Other avenues to be able to perform the Fisher Exact test were not explored, since the statistical test for bin testing would already provide the desired answers, regardless of a preceding overall test.

Therefore, a simulation approach has been used to test the above hypotheses.

3. Results

3.1. Yellow:number effect irrespective of window of correction

Table 1 shows that the number of SPADs is significantly higher for those approaches where the yellow signal was passed with a yellow:number aspect over 350 times and less than 550 times in the last 14 days. The bins with a frequency of 0 and 1–50 have the largest number of red aspect approaches and a SPAD percentage of respectively $2.00*10^{-3}$ and $1.55*10^{-3}$. The two significant bins have a SPAD percentage of $13.80*10^{-3}$ and $13.44*10^{-3}$, indicating not only a significant but also a large effect.

The absence of SPADs in the bin with frequency 251–350 is not in line with the hypothesis, but not surprising, given the relatively low number of red aspect approaches in this bin.

3.2. Yellow:number effect in combination with size of window of correction

Table 2 shows that the number of SPADs is significantly higher for those approaches where the yellow signal was passed with a yellow:number aspect over 350 times and less than 450 times in the previous 14 days and the window for correction was small. The SPAD percentage of $15.18*10^{-3}$ is similar to the SPAD percentage in the previous analysis for the bin with the same 14-day frequency. In the current sample, the bin with a frequency of zero was not included, leading to a higher average SPAD percentage, potentially contributing to why the bin with frequency 451–550 is not significant in this analysis, apart from a lower, albeit still high, SPAD percentage. The higher average SPAD percentage can also explain why the SPAD percentage of $1.16*10^{-3}$ in the bin with a frequency of 1–50 is now significantly lower, whilst in the previous analysis it was also lower than average but not statistically significant.

No SPADs were measured in the category of approaches with a large window for correction. Only three out of twenty SPADs occurred in the category of a medium window for correction, but it should be noted that the number of red aspect approaches was low for this category and a 14-day frequency above 250.

4. Conclusions and discussion

The results indicate a significant and large effect of exposure in the previous 14 days in combination with a small window for correction. This evidence supports the hypothesis that incorrect schema activation is a significant contributor in SPAD causation if there is a small window for correction.

A limitation of this research was that the exposure frequency was calculated by train series and not by train driver. It would have been preferable to measure how often the train driver had previously experienced the similar aspect, but information about the train driver was not disclosed for privacy reasons. Despite suboptimal data, this research was still possible because it focuses on relative changes. When a train series has a yellow:number frequency of 300 in the previous 14 days, it is likely that the specific train driver did not pass a yellow:number aspect in that location all 300 times. However, the train driver is likely to have seen a greater number of yellow:number aspects than in those cases where the frequency was only 100. The Netherlands has 28 work locations for drivers, with each location having certain work packages, including some variation in routes but also repetition of routes by the same drivers (Jacobs, 2016; Standplaats Available online, 2020; Slidestops Hoeveel trajecten rijdt een, 2352). Additionally, the authors analysed Dutch SPAD reports and frequently noticed train driver statements such as ‘usually in this location there is aspect xyz’, further supporting the notion that the Dutch train drivers indeed drive the same routes repeatedly.

Nonetheless, the research would be improved by replication using driver data. This would also give more insight into how often an employee actually needs to be exposed to a certain situation for incorrect schema activation and execution to occur, which would give better guidelines for SPAD prevention.

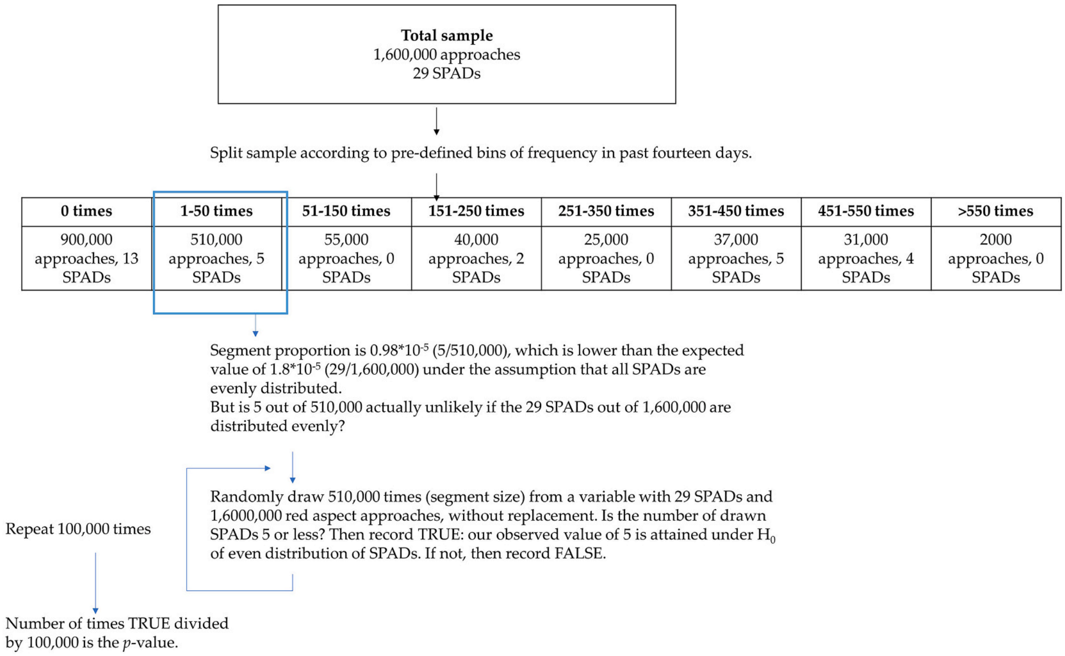

Fig. 5. The p-value is calculated per segment by comparing the observed number of SPADs with the number of expected SPADs if the SPADs are distributed evenly.

Table 1
SPAD percentage is significantly higher for a frequency of yellow:number aspects between 351 and 550 times in previous 14 days.

# times in previous 14 days	0	1-50	51-150	151-250	251-350	351-450	451-550	>550
# SPADs	13	5	0	2	0	5	4	0
# RAA	649,738	323,313	48,057	31,053	20,082	36,227	29,752	1,443
% * 10^{-3}	2.00	1.55	0	6.44	0	13.80	13.44	0
p-value	N.S.	N.S.	N.S.	N.S.	N.S.	0.002	0.006	N.S.

Table 2
SPAD percentage is significantly higher for a frequency of yellow:number aspects between 351 and 450 times in previous 14 days and a small window for correction.

Window for correction:	Frequency:	1-50	51-150	151-250	251-350	351-450	451-550	>550
Large (<0.6 m/s^2)	#SPADs #RAA %* 10^{-3} p-value	0 544,037 0 N.S.	0 103,675 0 N.S.	0 45,036 0 N.S.	0 38,098 0 N.S.	0 29,311 0 N.S.	0 17,002 0 N.S.	0 351 0 N.S.
Medium (0.6-0.8m/s^2)	#SPADs #RAA %* 10^{-3} p-value	2 245,326 0.81 N.S.	0 30,958 0 N.S.	0 13,325 0 N.S.	0 7,996 0 N.S.	0 13,654 0 N.S.	1 8,063 12.40 N.S.	0 211 0 N.S.
Small (>0.8m/s^2)	#SPADs #RAA %* 10^{-3} p-value	4 345,537 1.16 0.001	0 61,080 0 N.S.	3 40,157 7.47 N.S.	1 24,695 4.05 N.S.	6 39,520 15.18 <0.001	3 34,545 8.68 N.S.	0 1,928 0 N.S.

Another avenue for future research is including the distribution between different types of exposure. Our independent variable was how often the yellow:number aspect was present in the past fourteen days, for a location where there was currently a yellow aspect. This gave no information about how often that specific signal was yellow (followed by a red aspect) in the past fourteen days. In general, the yellow:number aspect as part of a scheduled approach are more frequently present at a given location than the unscheduled approach with a yellow. This gave rise to the labeling of yellow:number as the 'standard' approach and the 'yellow' approach as the 'deviating approach'. It was however not explicitly taken into account how often train drivers are exposed to the deviating approach in addition to the exposure to the standard approach. When this information is taken into account in future research, it could also be relevant to include the sequence of exposure. A train driver might be exposed to the standard approach during 90 approaches and to the deviating approach during 10 approaching, but does it matter if the train driver is exposed to all 90 in a row before being exposed to a deviating approach versus for example 10 standard approaches followed by 1 deviating approach followed by a number of standard approaches and then a deviating approach again? Especially when including the latter variable, data on train driver level rather than on train series level is recommended.

Based on the results of the current study, locations with a higher probability of a SPAD can be identified in Dutch rail if the following questions are all answered with 'yes':

- Are there situations where yellow:number-yellow-red aspect sequences can also be yellow-red in the same location?
- Is the yellow:number aspect frequently present for that location?
- Is there a small window for correction, i.e., is there a large difference between the required deceleration during the yellow:number-yellow versus the yellow-red approach?[1]
- Are there no other SPAD prevention mechanisms present such as the auditory warning system?

Interventions to prevent SPADs can be aimed at preventing wrong schema activation, increasing the opportunity for self-correction and/or implementing intervention mechanisms. To prevent the possibility of the wrong schema activation leading to an error, the infrastructure design can be improved via adjusted signal placement and/or track speed to make sure there are no locations where the yellow:number-yellow-red aspect sequence can also be yellow-red in the same location ánd there is a small window for correction. When this is not possible for a given location, the probability of wrong schema activation can be reduced by a) decreasing the frequency of yellow:number aspects or b) increasing the dissimilarity between the yellow and yellow:number signal aspects. To increase the opportunity for self-correction when wrong schema activation cannot be prevented, a) the window of correction can be increased, b) measures can be taken to increase the probability that the red aspect is perceived from afar (e.g. visibility and line of sight), and c) the braking power of the trains can be increased. Another option is employing other SPAD prevention methods such as an auditory warning system or automatic intervention by technical systems.

The Dutch Railways started implementing the Orbit warning system in 2018. Initial results of in-company research projects within ProRail and NS have shown indications that Orbit has helped to prevent SPADs, although future research is needed to provide evidence that an auditory warning system such as Orbit can also prevent SPADs in which wrong schema activation occurred. Nonetheless, understanding the causes of unsuitable deceleration behaviour remains important even when Orbit is also a useful safety barrier in preventing these types of SPADs. The system simply does not cover all red aspect approaches because it is not implemented in all trains yet and is not designed to cover all signals. Technical failures are also possible. We advocate the presence of multiple safety barriers, including supporting the train driver to drive as desired by improving the infrastructure design.

Another reason that this research remains relevant despite technical advances is that it touches upon a larger topic, namely the need to take previous exposure into account for optimal task design. If (visually) similar situations often require one type of behaviour and sometimes require different behaviour, then the occurrence of an error should be considered. Employees can correct their own initial error if they perceive a clear signal (such as a red aspect) in time. It is plausible that this effect does not only apply to the interaction between the train driver and signal aspects along the tracks, but also during the interaction between the train driver and on-board systems.

Funding

This research was funded by ProRail, the Dutch rail infrastructure manager.

[1] ProRail has requirements for minimum distance between signals per speed and speed reduction. In the case of a zero percent slope and a track speed of 80 km/h, the size of the window for correction is 0.71 m/s^2 for the shortest possible distance in combination with a yellow:6 aspect. When the automatic train protection system ATB-EG is present in the tracks, the signal distance must be larger and the size of the window for correction is 0.62 for the shortest possible distance in combination with a yellow:6 aspect. The size of the window for correction is 0.51 m/s^2 or 0.46 m/s^2 for the shortest possible distance in combination with a yellow:4 aspect without and with ATB-EG presence respectively. At a zero percent slope and a track speed of 60 km/h, the size of the window for correction is 0.45 m/s^2 or 0.39 m/s^2 for the shortest possible distance in combination with a yellow:4 aspect without and with ATB-EG presence respectively. Thus, at a track speed of 80 km/h, the size of the window for correction is either large ($<$0.6 m/s^2) or medium (0.6–0.8 m/s^2) and at a track speed speeds of 60 km/h the size of the window for correction is always large ($<$0.6 m/s^2).

Acknowledgements

We would like to thank Heidi van Spaandonk, manager Safety National Advice and Data at ProRail, for advocating this research and for her project management advice, and Rosa Sterkenburg, Data Scientist at ProRail, for her extensive work using Spark to turn the raw data in the necessary format.

Appendix

R code for statistical test:

```
## Statistical testing: get p-value for 1 segment by comparing observed with expected
# below values for frequency bin 351-450 times in Error! Reference source not found.
SPADs<-13+5+0+2+0+5+4+0
RAA<-649738+323313+48057+31053+20082+36227+29752+1443 -SPADs
tot<-SPADs+RAA
values<-c(rep(1,SPADs), rep(0,RAA))
# example values in subset
SPADs_subset<-5
n_subset<-36227
# Prepare for 100000 runs
reps<-100000
result<-logical(length=reps)
# Check side: greater or lesser than; if greater than expected:
if(SPADs_subset>(SPADs/tot*n_subset)){
 for(i in 1:reps){
   # draw random sample of subset size without replacement and check if as many or more values in drawn as in measured
  result[i]<-sum(sample(values, n_subset, replace = F)==1)>=SPADs_subset
 }
}
# Check side: greater or lesser than: if lesser than expected:
if(SPADs_subset<(SPADs/tot*n_subset)){
 for(i in 1:reps){
   # draw random sample of subset size without replacement and check if as little or less values in drawn as in measured
   result[i]<-sum(sample(values, n_subset, replace = F)==1)<=SPADs_subset
 }
}
p<-sum(result)/reps
p
```

References

Balfe, N., Doyle, K., 2021. Delving deeper: applying human factors analysis to identify factors contributing to railway incidents. In: Proceedings of the 7th International Human Factors Virtual Rail Conference 2021. Virtual.

Bartlett, F., 1932. Remembering: A Study in Experimental and Social Psychology. Cambridge University Press, Cambridge, UK.

Baysari, M.T., McIntosh, A.S., Wilson, J.R., 2008. Understanding the human factors contribution to railway accidents and incidents in Australia. Accid. Anal. Prev. 40, 1750–1757.

BNR Webredactie CEO ProRail, 2020. 30 Procent Groei in 2030.

Burggraaf, J., Groeneweg, J., Sillem, S., van Gelder, P., 2021. What employees do today because of their experience yesterday: how incidental learning influences train driver behavior and safety margins (A big data analysis). Saf. Now. 7, 2.

Gazzaniga, M., Ivry, R., Mangun, G., 2009. Memory and brain. In: Cognitive Neuroscience: the Biology of the Mind. W.W. Norton & Company, pp. 324–356.

Gertman, D.I., Blackman, H.S., Marble, J.L., Smith, C., Boring, R.L., O'Reilly, P., 2004. The SPAR H human reliability analysis method. Am. Nucl. Soc. 4th Int. Top. Meet. Nucl. Plant Instrumentation, Control Hum. Mach. Interface Technol. 17–24.

Gibson, W.H., Willet, J., Lewis, G., Harrison, C., 2017. Exploring the limits of train driver reliability. In: Proceedings of the Sixth International Human Factors Rail Conference. London, England.

Gibson, H., Roels, R., Harrison, C., Kohli, R., 2021. Underlying causes of signals passed at danger - looking at a spike. In: Proceedings of the 7th International Human Factors Virtual Rail Conference, p. 2021.

GWW Grotere capaciteit spoorgoederenvervoer in 2030 en nog duurzaam ook! Available online: gww-bouw.nl/artikel/grotere-capaciteit-spoorgoederenvervoer-in-2030-en-nog-duurzaam-ook (accessed on Apr 28, 2021).

Hamilton, W.I., Clarke, T., 2005. Driver performance modelling and its practical application to railway safety. Appl. Ergon. 36, 661–670.

Harrison, C., Stow, J., Ge, X., Gregory, J., Gibson, H., 2021. Using the Red Aspect Approaches to Signals (RAATS) tool to better understand the human relaibility associated with SPADs. In: Proceedings of the 7th International Human Factors Virtual Rail Conference, p. 2021.

Jacobs, I., 2016. Variatie Bieden in Routes NS-personeel Wordt Lastiger. OVPRO. https://www.ovpro.nl/special/2016/11/01/rondje-om-de-kerk-ns-lijkt-onvermijdelijk/?gdpr=accept.

Kyriakidis, M., Simanjuntak, S., Singh, S., Majumdar, A., 2019. The indirect costs assessment of railway incidents and their relationship to human error - the case of Signals Passed at Danger. J. Rail Transp. Plan. Manag. 9, 34–45.

Lawton, R., Ward, N.J., 2005. A systems analysis of the Ladbroke Grove rail crash. Accid. Anal. Prev. 37, 235–244.

Madigan, R., Golightly, D., Madders, R., 2016. Application of human factors analysis and classification system (HFACS) to UK rail safety of the line incidents. Accid. Anal. Prev. 97, 122–131.

Ministerie van Infrastructuur en Waterstaat *Basisrapportage, tevens dertiende voortgangsrapportage European Rail Traffic Management System*, 2020. ERTMS.

Ministry of Infrastructure and the Environment *Railway Map ERTMS Version 2.0 - State Of Play Regarding Research in the Exploratory Phase*, 2013.

Naweed, A., Hockey, G.R.J., Clarke, S.D., 2013. Designing simulator tools for rail research: the case study of a train driving microworld. Appl. Ergon. 44, 445–454.

Naweed, A., Bowditch, L., Chapman, J., Balfe, N., Dorrian, J., 2019. System Precursors to Signals Passed at Danger (SPADs): an Exploratory Comparison of SPAD History and Rail Environment, pp. 1–9.

Neumann, O., 1984. Automatic processing: a review of recent findings and a plea for an old theory. In: Prinz, W., Sanders, A.F. (Eds.), Cognition and Motor Processes. Springer Berlin Heidelberg, Berlin, Heidelberg, Germany, pp. 255–293.

Norman, D.A., 1981. Categorization of action slips. Psychol. Rev. 88, 1–15.

Norman, D.A., Shallice, T., 1986. Attention to action. In: Consciousness and Self-Regulation. Springer US, Boston, MA, pp. 1–18.

Ontwerpvoorschrift, ProRail, 2020. Plaatsing en toepassing van lichtseinen.

Piaget, J., Cook, M.T., 1952. The Origins of Intelligence in Children. International University Press, New York, NY, USA.

Punzet, L., Pignata, S., Rose, J., 2018. Error types and potential mitigation strategies in Signal Passed at Danger (SPAD) events in an Australian rail organisation. Saf. Sci. 110, 89–99.

Singer, M.H., Lappin, J.S., 1976. Similarity: its definition and effect on the visual analysis of complex displays. Percept. Psychophys. 19, 405–411.

Slidestops Hoeveel trajecten rijdt een machinist eigenlijk?. Available online: https://community.ns.nl/off-topic-32/hoeveel-trajecten-rijdt-een-machinist-eigenlijk-36445?postid=235209#post235209. (Accessed 2 July 2021).

Smith, P., Majumdar, A., Ochieng, W.Y., 2012. An overview of lessons learnt from ERTMS implementation in European railways. J. Rail Transp. Plan. Manag. 2, 79–87.

Standplaats. Available online: https://wiki.ovinnederland.nl/wiki/Standplaats. (Accessed 3 December 2020).

Stanton, N.A., Walker, G.H., 2011. Exploring the psychological factors involved in the Ladbroke Grove rail accident. Accid. Anal. Prev. 43, 1117–1127.

Summerfield, C., Egner, T., 2009. Expectation (and attention) in visual cognition. Trends Cognit. Sci. 13, 403–409.

Wickens, C.D., Lee, J., Liu, Y.D., Gordon-Becker, S., 2003. Introduction to Human Factors Engineering, second ed. Pearson Education (US), Upper Saddle River, NJ, USA.

Wiegmann, D.A., Shappell, S.A., 2003. A Human Error Approach to Aviation Accident Analysis: the Human Factors Analysis and Classification System. CRC Press LLC, Abingdon, UK.

Wilson, J.R., Norris, B.J., 2005. Rail human factors: past, present and future. Appl. Ergon. 36, 649–660.