CrossMark

ORIGINAL ARTICLE

Toward quantifying metrics for rail-system resilience: identification and analysis of performance weak resilience signals

Anouk de Regt[1] · Aron Wolf Siegel[1] · Jan Maarten Schraagen[1,2]

Abstract This paper aims to enhance tangibility of the resilience engineering concept by facilitating understanding and operationalization of weak resilience signals (WRSs) in the rail sector. Within complex socio-technical systems, accidents can be seen as unwanted outcomes emerging from uncontrolled sources of entropy (functional resonance). Various theoretical models exist to determine the variability of system interactions, the resilience state and the organization's intrinsic abilities to reorganize and manage their functioning and adaptive capacity to cope with unexpected and unforeseen disruptions. However, operationalizing and measuring concrete and reliable manifestations of resilience and assessing their impact at a system level have proved to be a challenge. A multi-method, ethnographic observation and resilience questionnaire, were used to determine resilience baseline conditions at an operational rail traffic control post. This paper describes the development, implementation and initial validation of WRSs identified and modeled around a 'performance system boundary.' In addition, a WRS analysis function is introduced to interpret underlying factors of the performance WRSs and serves as a method to reveal potential sources of future resonance that could comprise system resilience. Results indicate that performance WRSs can successfully be implemented to accentuate relative deviations from resilience baseline conditions. A WRS analysis function can help to interpret these divergences and could be used to reveal (creeping) change processes and unnoticed initiating events that facilitate emergence that degrades rail-system resilience. Establishing relevant change signals in advance can contribute to anticipation and awareness, enhance organizational learning and stimulate resilient courses of action and adaptive behavior that ensures rail operation reliability.

Keywords Resilience · Ethnographic observation · Railway signaling · Weak resilience signal · WRS · WRS analysis function · Resilience state model for railway systems

1 Introduction

We currently live in an increasingly tightly coupled and interactively complex world in which unpredictable events are omnipresent, and the velocity with which unanticipated events can amplify into unwanted outcomes is continually increasing (Weick and Sutcliffe 2001). Within this setting, the railway industry is broadly recognized as an example of a safety-critical and complex socio-technical system (e.g., Wilson et al. 2007; Belmonte et al. 2011). To maintain control, enhance efficiency and improve safe operations in the rail industry, a rise in automation (Wilson and Norris 2005), standardization and strict adherence to protocols and predefined timetables (off-line time-tabling; Goverde and Odijk 2002; Hansen 2010) has been notable over the years. This optimization rationale enabled the European railway to become an ultra-safe system (one accident per one million events; Amalberti 2001; European Railway Agency 2014). This, combined with the search for sustainable transport solutions, induced political focus on rail transportation throughout Europe (Ferreira et al. 2011).

✉ Aron Wolf Siegel
A.W.Siegel@UTwente.nl

1 University of Twente/BMS, P.O. Box 217, 7500, AE, Enschede, The Netherlands

2 TNO Earth, Life, and Social Sciences, P.O. Box 23, 3769, ZG, Soesterberg, The Netherlands

However, to meet and maintain the high levels of performance (e.g., punctuality, capacity and safety) that are required to realize this potential, linear and additive optimization solutions (e.g., more rules and regulations) may prove to be insufficient (e.g., Bieder and Bourrier 2013). Unwanted outcomes in ultra-safe complex socio-technical systems 'emerge' from a combination of unanticipated, nonlinear relationships between constituent parts of the system that can arise under dynamic operating conditions (Leveson 2004; Dekker et al. 2008). This causes the system to contain hidden fragilities with respect to rare and relatively unpredictable perturbations, making the system robust yet fragile (RYF; Doyle et al. 2005). The railway system thus faces the challenge of finding alternative methods to enhance performance and outmaneuver (confusing) system complexity (de Carvalho 2011).

The resilience (systems) approach is considered to be the next step (e.g., Qureshi 2008) and has become, arguably, the dominant paradigm in the study of complex socio-technical systems (Underwood and Waterson 2013). Resilience engineering can be defined as a proactive approach concerned with enhancing organizations' intrinsic abilities to reorganize and manage their functioning and adaptive capacity prior to, during or following events, so that the system can sustain the required level of operations under both expected and unexpected conditions (Woods and Branlat 2010; Hollnagel 2014). Resilience should thus be seen as an emergent property originating from what an organization does, rather than what an organization has, emphasizing function over structure and ability over capacity (Hollnagel 2004). Different theoretical models are available and have been used over the years to describe the resilient state of a system (e.g., the ball and cup model: Scheffer et al. 1993; safe operating envelope: Rasmussen 1997; stress–strain (S–S) model: Woods and Wreathall 2008; resilience analysis grid (RAG): Hollnagel 2011; functional resonance analysis method (FRAM): Hollnagel 2012). Among resilience researchers, there is general consensus that people are the primary source of resilience (e.g., Woods et al. 2007). In accordance, providing techniques and system designs that help people and organizations cope with complexity might thus be one method to enhance system resilience. However, without a clear understanding of what manifestations of resilience look like (Back et al. 2008), it will be difficult to identify such manifestations in practice and quantify the theoretical models developed, creating a research–practice gap (Underwood and Waterson 2013). This is especially true when focusing on quantifying resilience for infrastructural systems, in which the current quantification methods used (e.g., graph theory: Berche et al. 2009; fuzzy interference: Heaslip et al. 2010) emanate from other well-established and well-elaborated methodological frameworks but as

such are not fully capable of capturing the underlying interrelations of system modules (Tamvakis and Xenidis 2013). Research aimed at operationalizing theoretical resilience models and prospective analysis frameworks for quantifying resilience of infrastructure systems is required (e.g., Madni and Jackson 2009).

Siegel and Schraagen (2014) contributed to diminishing the research–practice gap by developing a so-called 'resilience state model' for railway systems. This model is based on Rasmussen's (1997) safe operating envelope, the stress–strain model described by Woods and Wreathall (2008), and adheres to the notion that knowledge and error flow from the same mental sources (Mach 1905; Hollnagel 2012). Siegel and Schraagen (2014) adapted the three (relative) system boundaries defined by Rasmussen (1997) (performance, economy and workload) to describe and explain the various (external) pressures—safety, performance (capacity and punctuality) and workload—that affect the operating state of a railway system. In addition, a depth dimension was added to the model that enables differentiation between internal changes that keep the system in a resilient state or have it move toward brittleness. The stress–strain model (Woods and Wreathall 2008) characterizes the properties of an organization as an adaptive system by using an analogy from materials engineering which focuses on the relationship between the external and varying demand on a mechanical structure (stress), and how the structure stretches in response (strain). Siegel and Schraagen (2014) use changes in the linear relation between stress and strain (i.e., Young's modulus slope) to model quantifiable rail weak resilience signals (WRSs). A WRS indicates a change in the system's operating state and is defined by measuring properties in the base capacity region of the system that signals changes of properties in the extra adaptive region of the system. In this model, the base capacity reflects the 'normal' functioning response of the system to external events. The extra adaptive capacity reflects the potential discrepancy between adaptive system responses and external demands that challenge or fall outside the boundaries of the base operating capacity (Woods et al. 2014). In other words, WRSs represent uncertain snippets of information, hidden within the 'normal' system variability, which could be used as situated indicators to signal potential change processes in the organizations' resilience level. The 'resilience state model' for railway systems is thus described as a framework for generic quantifiable modeling of rail WRSs around three (relative) system boundaries (workload, performance and safety; Siegel and Schraagen 2014). Previous research indicated that changes around the workload boundary could be successfully measured, identified and used to quantify workload WRSs (Siegel and Schraagen 2014). In this paper, we extend implementation of the WRS

framework to the performance boundary. Since within rail systems the quality of performance is, to a large extent, based on time-related key performance indicators, specific methods to measure and quantify changes in punctuality and rail capacity need to be defined.

To develop operational parameters that identify changes around the performance boundary, baseline conditions (acceptable levels of performance) of the current operating state should be established where deviations can be measured against. Examining the current operating state will enhance insight into and understanding of how WRSs arise and support the interpretation of the WRSs indicated through the WRS framework. In addition, focusing on different elements that comprise or serve as alternative and additional performance indicators (i.e., analysis functions) might help to further enhance interpretation, understanding and analysis of rail WRSs indicated through the framework. To investigate the factors influencing the operating state, this paper adheres to Hollnagel's (2009) notion that management of uncertainty and system variability in the (real-time) operation is built around four main system capabilities defining resilience: responding to the actual (knowing what to do), monitoring the critical (knowing what to look for), anticipating the potential (knowing what to expect) and learning from the factual (knowing what has happened).

The overall aim of this research is to enhance tangibility of resilience by facilitating understanding and operationalization of rail weak resilience signals. The research question is threefold: (1) How can the baseline conditions that comprise the current resilience operating state of the socio-technical rail system be determined? (2) How can the resilience state model for railway systems be further developed to enable quantification and operationalization of performance WRSs? and (3) Can a WRS analysis function enhance interpreting and understanding of (creeping) factors that underlie performance WRSs?

The rest of this paper is organized as follows: the multi-method research approach will be described, and operational WRS parameters will be introduced. Resilience baseline conditions will be identified, and quantification of WRSs on the punctuality boundary will be explicated upon. In conclusion, the main points and results will be discussed and an outlook on future research will be provided.

2 Method

A multi-method approach was used to acquire data and knowledge about the current operating state of the system (i.e., identify performance indicators). Influenced by the interdisciplinary research fields of human–computer interaction and computer-supported cooperative work (e.g.,

Herrmann et al. 2004; Millen 2000), an ethnographically informed method was adopted to construct adequate understanding of the working environment, discover exceptional and beneficial user behavior and provide additional insights into social and organizational phenomena.

2.1 Research setting

The research took place at a rail control post in Zwolle. This rail control post, one out of 13, handles rail stations and train traffic in the northeastern part of the Netherlands. To bolster rail operation reliability and efficiency, the setup of the control room (Fig. 1) consists of eight workstations divided into two main rail corridors: south–north, called corridor North (four workstations), and west–east, called corridor East (three workstations). A corridor is defined as a large area of rail infrastructure that consists of adjacent rail sections from various (interdependent) workstations. Within this setup, train dispatchers work in small 'corridor' teams consisting of approximately 3–4 people, one planner (executive logistics focus) and 2–3 train dispatchers (safety focus), who rotate in positions. The one remaining workstation is exclusively utilized during calamities and is readily available to either corridor as a means to provide added resources and capacity.

The population consisted of a convenience sample of 25 rail dispatchers (five females, 20 males) working at least 24 h a week. Because both senior as well as junior dispatchers were included in the sample, years of experience in train dispatching ranged from 2 to 37 years ($M = 18.5$, $SD = 10.55$). In addition, age varied between 29 and 65 years ($M = 47$, $SD = 8.72$).

2.2 Resilience observations

The observations took place over a period of 2 non-contiguous weeks. During the first week, three researchers pulled rotating shifts of 8 h to fully cover the rail dispatcher's proceedings during day and night shifts (24/7). In the second week, two researchers observed the domain practitioners between 7:00 AM and 9:00 PM, logging a total of 266 h of observation. Since the presence of people other than the rail dispatchers is not uncommon in the control room, disruption of the natural setting was minimized. This non-obtrusiveness allowed the researchers to roam around freely to observe and interview the rail dispatchers as opportunities presented themselves. In order to guide this process and enhance the efficiency, outcomes of real-time workload measurements were used to focus on situations in which train dispatchers experienced increased workload. Workload measurements were conducted using an automated version of the Integrated Workload Scale
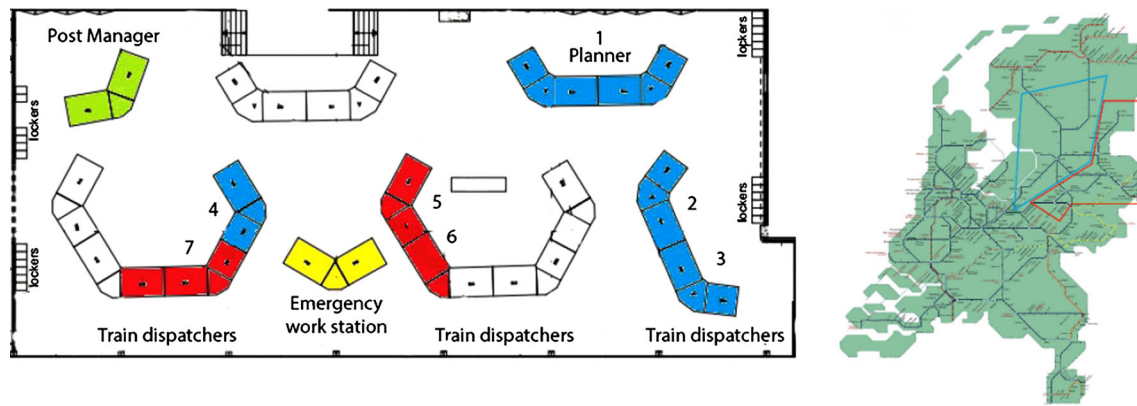
Fig. 1 Corridor North (1, 2, 3, 4) and corridor East (5, 6, 7) and corresponding rail trajectories
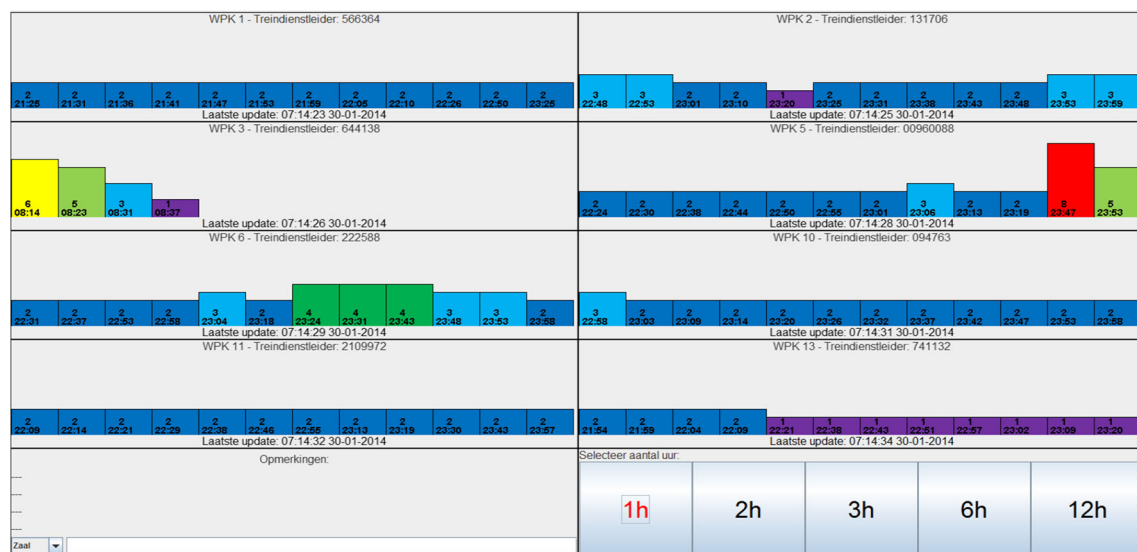

Fig. 2 Monitoring interface for the observer (eight workstations with resolution time setting)

(IWS; Pickup et al. 2005), a unidimensional self-report tool consisting of a 9-point scale in which 1 = not demanding and 9 = work too demanding, that enabled numerical entering of rail dispatchers' subjective workload at 5-min intervals. An interface was developed to display the 5-min interval entering of the subjective workload for all individual rail dispatchers and all individual workstations in real time (Fig. 2). To distinguish between different rail dispatchers who worked at the same working station, a unique but anonymous numerical code had to be entered at the beginning of a rail dispatcher's shift. The purpose of the interface was to enhance rapid identification of (relatively) busy work situations. To enable this, the height of the bar charts was adjusted according to the 5-min interval scores and different (rainbow) colors were chosen to visualize the IWS scores (with 1 = purple and 9 = dark red). The general configuration of the interface was set to display the current hour; however, time slots of 2, 3, 6 and 12 h were also available.

Although the observations as well as the conversations were mainly centered around the dispatchers' current actions, the work in general and prior situations (that either turned out unexpectedly good or resulted in unwanted outcomes and safety issues) was also discussed with dispatchers retrospectively. The purpose of this was twofold: (1) to enhance understanding and (2) to fill the (potential) observational research gap which occurs due to the fact that the European railway has become an ultra-safe system (Amalberti 2001). Since in an ultra-safe system on average only one accident occurs per a million events, the chances of us observing such an event during the observation period are close to zero. It is, however, of utmost importance to know how the dispatchers would handle, or have handled, such situations. Therefore, within the context of the current

as well as the prior situations, particular attention was placed on capturing concrete examples in which the complexity of the environment, in either routine or disturbed situations, required resilient behavior. The Resilience Markers Framework by Furniss et al. (2011), which is used to systematically observe concrete manifestations of resilience at different levels of granularity within and across domains, was used to document the cognitive, collaborative strategies and facilitating activities used to handle situation demands. This framework distinguishes three levels: (1) resilience markers, (2) resilience strategies and (3) resilience observations. Attributes and an example of the different levels are shown in Table 1.

2.3 Weak resilience signals

To resolve potential resilience gaps that surface from the base activity (i.e., through observations) and reveal other issues and sources of future resonance that could comprise system resilience in the long term, operationalization and implementation of WRSs were conducted. To measure workload WRSs, Siegel and Schraagen (2014) introduced a new metric, stretch, which can be defined as an objective or a subjective system reaction to an external (cluster) event. An objective stretch is used to identify an absolute workload growth and is comprised of two main components: task complexity and events. Task complexity is composed of real-time technical system measurement and readily available (log) data of rail dispatchers' main job requirements: monitoring of rail movement, performance of plan mutations, execution of manual actions and communication activities. Events are defined as external events, i.e., not controlled by the operators themselves, that can influence the operator's task load (e.g., section and switch disruptions, rail track maintenance data, number of delayed trains and the number of phone calls). The total number of events is calculated and measured in 5-min intervals. This value is normalized between 1 and 2 and multiplied with the task complexity resulting in the external task load (XTL). The subjective stretch is the human perception of the system's strain and embodies the (cumulative) workload effort during a period of time in which IWS scores deviate from an IWS baseline. In this context, an IWS baseline is defined as the steady-state IWS rating before and after a disruption occurs (i.e., the IWS rating during scheduled rail movements). Since we were looking for relative changes in workload experienced by individual rail dispatchers, it was decided not to weigh the IWS scores to account for subjective variability due to competence. The time span of the objective and subjective stretch can differ if the activity in the system started earlier or ended later than the workload shift perceived by the rail dispatcher. Therefore, the start time and end time of a stretch are adjusted to the first XTL-minimum moment before the IWS rising from, and after the IWS returning to, the IWS baseline. The ratio between the subjective and objective stretch is used to identify workload WRSs.

When a growing change of a stretch ratio is identified and the stretch values are larger than a predefined (threshold) value, a weak resilience signal (WRS) is generated. To indicate significant and relative changes when comparing two periods, the accumulated standard deviation (SD) of the stretch ratio in each period was used (for more information, see Siegel and Schraagen 2014). The workload WRSs were measured and used as a starting point and reference frame, to extend operationalization of the rail WRS framework to the performance boundary. To operationalize and utilize performance WRSs, identification and implementation requirements had to be established.

The operational parameters for the performance boundary are entirely based on technical system measurement and readily available (log) data. Since the rail capacity is generally stable over the year (e.g., due to predefined and optimized offline timetabling; Goverde and Odijk 2002; Hansen 2010), differences between the predefined timetable (scheduled planning) and the working timetable (real-time measurements of rail movements) were used to measure performance WRSs. The main focus was placed on delay development, and propagation, within the rail system and how this impacts the punctuality boundary (i.e., buffering capacity, flexibility, margin and tolerance; Woods and Cook 2006). Adhering to rail-dispatching guidelines concerning time lag and rail movement, a train was considered delayed if the deviation from the pre-defined timetable exceeded a 3-min threshold.

Table 1 Levels, attributes and an example of the hierarchy (Furniss et al. 2011)

Level	Generalizability	Quantity	Example
Markers	High	Low	Maximizing information extraction
Strategy	Moderate	Moderate	Creating an external cue
Observation	Low	High	A paperclip to bookmark a page in the procedure someone is following

By providing actionable attention cues, the WRSs will contribute to revealing eroding levels of (operational) system resilience. The attention cues are visualized by generating graphical representations (WRS graphs created in Excel). In addition, the WRSs serve as an objective method (i.e., based on technical system data measurements) to approach the resilience base capability levels observed.

To guide the process of selecting WRSs that need to be dealt with, analysis functions will be constructed. An analysis function serves as an alternated frame of reference that is based on other or additional performance indicators (i.e., besides the punctuality data). The aim of implementing an analysis function is to exclude the 'evident, known and obvious' causes of resonance, and shift attention to reveal 'hidden, unmarked or ignored' processes and incident precursors that could affect rail-system resilience.

2.4 Resilience Questionnaire

Concurrent with the second week of observations, a resilience questionnaire was distributed by e-mail among all rail dispatchers and the management ($N = 67$) as a cross-referential method to measure the (operational) resilience level within the rail control post. An online rather than on-paper survey method was chosen due to the shift roster. To boost the response rate, a reminder e-mail was sent the following week and a reminder message was placed in the organization's weekly newsletter.

The ADAPTER questionnaire (Analyzing and Developing Adaptability and Performance in Teams to Enhance Resilience; van der Beek and Schraagen 2015) was selected because the questionnaire is suited to diagnose team resilience requirements of safety critical jobs and can be administered within a relatively short time period. In ADAPTER, the four essential abilities of resilience (Hollnagel 2009) are supplemented with relation-oriented abilities such as leadership and (cross-boundary) cooperation, to operationalize the concept of team resilience. Although the questionnaire was already available in Dutch, it was decided to slightly change the wording of some questions to better fit the terminology used within the railway organization. In addition, an N/A category option was added, where appropriate, to ensure valid answers and avoid positive skewness of answer categories which were not applicable for our specific situation (e.g., the N/A category was added to questions relating to 'cooperation with other teams' since in rail control this often involves cooperating across organizational boundaries in which not all information about the other teams is known or available). To evaluate the ADAPTER results, we used the method of van der Beek and Schraagen (2015) to compute

descriptive statistics and reliability estimates for the whole sample.

3 Results

3.1 Real-time dispatching observations

The transportation planning within the Dutch railway system is a highly dynamic multifaceted process. Within this context, train dispatchers coordinate and manage the (conflicting) demands placed on track use and integrate multiple sources of information to conduct trade-off decisions and actions necessary (e.g., re-routing, re-ordering and re-timing of trains, tracks and signals) to maintain performance, regain control and mitigate potential threats. Especially in uncertain, time pressured and variable traffic situations, in which train dispatchers are pushed toward the limits of their regular operating (base) capacity and the adaptive capacity of the system is challenged (e.g., Woods et al. 2014), handling the situational demands proves to be a cognitively complex task. It is in those instances that resilient strategies and behaviors are required and boundary conditions of adaptive capacity, as well as localization of those boundaries, might be exposed (Woods and Cook 2006; Dekker 2011). For this reason, observations and description of resilient behavior were focused around high-pressure situations.

In the next section, one of the observed high-pressure situations will be delineated. This illustrative case provides insight into the concrete manifestations of resilient dispatcher performance, as well as subsequent vulnerabilities, and serves as a baseline measure for the (operational) resilience conditions currently present within the organization.

3.1.1 Example of a high-pressure situation: 'the hooligan case'

On 2-4-2014, a major disruptive event unfolded when soccer hooligans ignited fireworks and smoke bombs on a rail station platform and the mobile police unit was forced to intervene. As is standard procedure in such high-pressure situations, the emergency workplace was put into operation (as a means of reinforcement to handle performance variability, minimize timetable disruptions and mitigate the rapidly increasing delays). In these situations, the dispatcher responsible for the rail trajectory in which the disruption occurs focuses on the direct (short-term) actions involved with the disruption handling (i.e., quick responsive action to train and timetable delays directly resulting from the disturbance) while the 'emergency' dispatcher assists by taking over verbal communication

with other actors (i.e., telephone calls from train drivers) and (long-term) planning activities. As is the case in almost all high-pressure situations, the corridor team was unable to integrate and develop implications for this specific situation based on the full set of information held by all actors involved (Woolley et al. 2008). This due to the fact that no direct line of communication could be established between the rail control center and the commander, nor other members, of the mobile police unit. Therefore, the corridor team initially chooses to arrange disruption handling of delayed trains based on the incoming order of notifications in the system. This method of prioritizing (short-term) proved to be inadequate and even counterproductive in the long term due to escalating knock-on delays for connecting trains (i.e., working at cross-purposes). This process was noticed by the post manager, approximately 15 min after the incident occurred, who directly stressed the importance of developing and implementing an action plan to properly deal with this situation. To fill the information gap, rail dispatchers used the live camera feed from the station platforms (Fig. 3). By monitoring the police actions on scene, rail dispatchers were able to enhance their overall situation awareness. Concurrently, the internal communication channels/structure were optimized. Two corridor team members gathered behind the rail dispatcher's workstation (responsible for the rail section where the disturbance took place) in order to listen in on communication and look at the monitor displays to gain insight into train movement and the overall rail situation on the surrounding tracks. Subsequently, this information was shared with train dispatchers manning the neighboring rail sections. Implementation of the action plan resulted in highly selective rail movement in the disruptive rail section (e.g., prioritizing international trains), and gradual redirection of

Fig. 3 Live camera feed from the station platforms

stationary and delayed trains occupying adjacent rail tracks to the nearest available railway station.

3.1.2 Resilience behavior episodes

Administering the Resilience Markers Framework by Furniss et al. (2011), two resilience behavior episodes were distinguished for this specific situation. (1) Recognition of inappropriate situation handling and avoiding escalation of commitment. (2) Tailoring of existing artifacts to maximize information extraction.

Recognition of inappropriate situation handling and avoiding escalation of commitment (i.e., the tendency to continue a chosen course of action even when changing to a new course would be preferable; (Staw 1981)) were related to the strategy 'provision of feedback to enable error correction' (Blandford and Furniss 2006) and the broader marker of 'recognizing and responding to failure'. Although the recognition and notification of malfunctioning initiated the corrective actions necessary to manage the performance variability in this situation, the insight came rather late and was only noted by one actor (the post manager) within the corridor team. Although it could be argued that the post manager has a high level of experience and as such might outperform the operational competence skill level of the other corridor team members, the tasks of a post manager and a rail dispatcher are of a different nature. As such, the post manager's skills and experience do not translate one-to-one to the abilities and experience of the corridor team members. An alternative explanation could be that the post manager provided a fresh perspective which led to a broader set of actions. This situation exposed potential vulnerabilities (e.g., maintaining adequate situational overview and awareness in high-pressure demands, acknowledgment of inappropriate actions and or routines) which could influence learning and anticipation of future resonance and disruption handling. This notion was strengthened by irregularities observed in the levels of operational performance within and between dispatchers and situations. Similar prioritizing decisions could be observed with other dispatchers over different shifts (e.g., answering incoming phone calls rather than prioritizing timetable changes, which would have been more efficient).

Tailoring of existing artifacts to maximize information extraction can be related to the strategies 'prepare for future work' (Blandford and Furniss 2006) and 'cue creation in action' (Perin 2005), with the broader markers of 'preparation' and 'strategies that maximize information extraction' (Blandford and Furniss 2006). The awareness of (incoming) data limitations and the proactive steps taken at present (i.e., enhanced monitoring) increased the readiness to adequately respond to ongoing developments

(efficient management of the performance variability) and provided the opportunity to anticipate and prepare for future situational demands.

3.1.3 Weak resilience signals

The operational parameters for the workload WRS set by Siegel and Schraagen (2014) were used to determine if the hooligan case could indeed be labeled as a high-pressure situation and if the system detected it as such. A WRS can be defined looking at three features: a (relatively) long stretch duration, high (average) IWS or XTL scores and discrepancies in the stretch ratio. A graphical representation of the workload stretch measures was generated by plotting all the objective versus the subjective stretches of that day relative to an empirically drawn threshold line (i.e., the rounded sum of the means with one standard deviation above). Since the stretches in Fig. 4 are significantly correlated ($r = .94$, $p < .05$), the threshold line serves as a visual guide to optimize the selection of stretches that deserve attention, by serving as a WRS. In Fig. 4, the x-axis represents the stretch duration x IWS scores. The stretch duration is derived by the sum of total minutes a stretch occurred representing the 5-min time slots. In addition, the mean IWS score can be calculated for each stretch. The y-axis indicates the sum of technical system activity measured in a specific stretch, also taking into account the 5-min time slots.

From the graph, it becomes clear that most stretches that occurred on 2-4-2014 were small and do not exceed the boundaries of the safe operating envelope (Rasmussen 1997). Ad hoc analysis revealed that five workload stretches in Fig. 4 are caused by the same underlying (decompensation) event, the ignition of fireworks and smoke bombs on the rail tracks and station platform by soccer hooligans ('Hooligan Case' in Fig. 4). Looking at these five stretches in relation to the three WRS features, it is

evident that all stretches have a (rather) long stretch duration with increased mean IWS scores (circa 5–6, indicating moderate pressure to very busy). In addition to an increased IWS average, all stretches also contained 5-min periods rated with the three highest IWS scores (7 = extreme effort, 8 = struggling to keep up, and 9 = work too demanding). The stretches also have increased levels of technical system activity (i.e., due to telephony and manual re-routing quantities) and enlarged deviations in the stretch ratio (see stretch number 2). All in all, the hooligan case can indeed be classified as a high-pressure situation.

To validate and verify the performance WRS constructs, (log) data were examined to determine whether the decompensation event that unfolded during that day could also have been identified using performance WRS data methods. A spike in delay development was identified for trains in the 1700 series, indicating a segment of 36 trains traveling the same rail trajectory (Fig. 5). The upward slope could be explained by three 'hooligan trains' (all part of the 1700 series). Two trains suffered imminent, rapidly increasing timetable delays due to the fact that they could not leave the station as a direct result of hooligans and fireworks on the tracks. The third train was used by the riot police to forcefully transport soccer hooligans out of the station. In addition, knock-on delays occurred due to the fact that trains retained from departure occupied the rail platforms. This induced red rail signals (an increase of 9.2 % above average) for connecting trains, forcing trains to wait on rail tracks surrounding the station.

3.1.4 WRS analysis function

The successful identification of the known hooligan event (i.e., it was observed during the ethnographic study) contributes to verification of the WRS method. Deviations from normal operational baseline periods, defined as the

Fig. 4 Workload stretch graph 2-4-2014 with hooligan case stretches
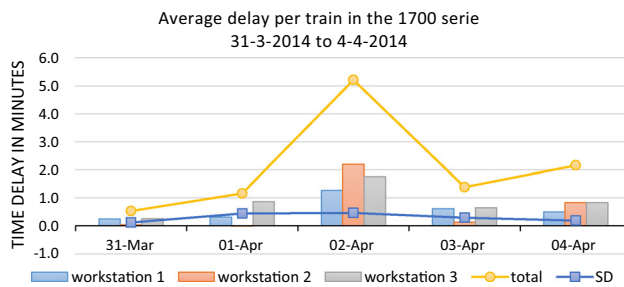
Hooligan case		
Stretch number	Mean IWS score	Stretch duration
1	6.5	00:55
2	4.7	01:35
3	4.8	01:45
4	4.8	01:55
5	6.0	01:35

Fig. 5 Performance WRS graph 31-3-2014–4-4-2014

Fig. 6 WRS analysis function applied on delay data from the 1700 series

steady state of a rail control post in which rail movements occur as planned without any intervention, could be established on the workload as well as the punctuality boundary. However, signaling of the hooligan event does not immediately create insight and understanding into the unknown variables in the normal performance variability that could indicate potential creeping sources of future resonance that may underlie the incident. Relatively long-lasting disruptive events with a big impact factor (i.e., affecting multiple trains and dispatchers) are likely to gain attention among actors in the system even without WRS indications. However, when such an event is already known, attention is needlessly diverted which may result in obscuring other unidentified potential factors that influence the resilience state. To enhance the organization's feedback control loop (Doyle et al. 2013) and increase the understanding, tracking and anticipation of potential sources of future resonance and or the impact factors of the different WRSs indicated by the framework, implementation of WRS analysis functions is proposed. An analysis function is described as an alternated frame of reference, based on other or additional performance indicators, which guides the process of selecting WRSs that need to be dealt with. The aim of this analysis function is to exclude the 'evident, known and obvious' causes of resonance, and attempt to shift attention and reveal 'hidden, unknown or ignored' processes that could affect rail-system resilience. To demonstrate the concept and implementation of this principle, a punctuality WRS analysis function was established for the hooligan scenario which will be described in more detail in the next section. It is important to note, however, that the use of analysis functions is not limited to high-pressure situations. Analysis functions are equally applicable to and well suited to uncover (creeping) incident precursors in routine situations.

Implementation of the WRS analysis function allows the frame of reference for the punctuality boundary to be manually altered by excluding trains with exorbitant delays due to well-known escalation events (i.e., the hooligan trains) and by comparing the real-time delay measurements to specified base line conditions (performance indicators;
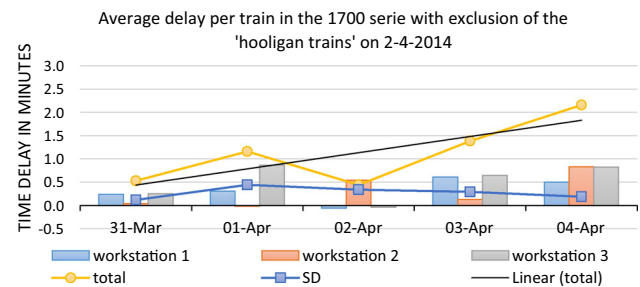
train series and specified dates). To test the applicability of the WRS analysis function method, the delay data from the second week of observations were re-examined (Fig. 6).

The three hooligan trains, which caused the exorbitant delays, were excluded from the analysis. Ad hoc analysis revealed an upward trend in delay development for the 1700 series. It could be argued that an average delay development increase of 1.7 min (102 s) per train does not exceed the predefined organizational threshold of ≥ 3 min delay and, as such, does not require further investigation. However, it could be beneficial to examine whether specific trains in this series contribute invariably to this delay development and whether this upward trend continues over time (e.g., the consecutive days or weeks). In addition, the time delays may impact the time buffers built-in on the pre-defined timetable and as such influence the rail dispatcher's workload. Such information could aid in forestalling and anticipating future resonance emerging from 'seemingly insignificant' (creeping) change patterns and might even identify commonalities in the operating state preceding well-known events.

WRSs and WRS analysis functions should be created to (visually) support the train dispatcher's comprehension of the current operating state and resilience status and to enhance prediction of possible incidents and accidents in the future by guiding attention to aspects that deserve further analysis. They provide a means to an end and will not in themselves present an integrated approach to improve the resilience or related aspects of the system. In other words, rather than directing the domain practitioners along a defined path, exploratory content that allows for comparison between data is provided.

3.2 Resilience questionnaire

The response rate to the questionnaire was calculated according to the American Association for Public Opinion Research (2015) RR1 definition. Of the 67 employees contacted, one person no longer worked for the company and a second person abstained due to prolonged absence.

Table 2 ADAPTER questionnaire; descriptive statistics and reliability coefficients

ADAPTER scales	n items	Mdn	SD	Cronbach's alpha (α)
Monitoring	13	3.69	.75	.95
Responding	12	3.46	.75	.90
Anticipating	8	3.25	.77	.68
Learning	9	3.17	.43	.70
Shared transformational leadership	14	3.14	.60	.91
Cooperation with other teams	12	2.54	.65	.71

$N = 22$. Total scale team resilience: $\alpha = .91$

This resulted in a RR1: 22/65 = 34 %, which is acceptable for online surveys (Nulty 2008). The sample demographics were as follows: 16 rail dispatchers, three managers and three front office employees. In total, two females and 20 males answered the resilience questionnaire. Results from the ADAPTER questionnaire (Table 2) are consistent with the resilience baseline conditions and current operating state of the system ascertained during the observations in that the domain practitioners rated the resilience constructs monitoring and responding higher than the resilience constructs anticipating and learning. It should be noted, however, that the scores for anticipating (Mdn = 3.25) and learning (Mdn = 3.17) fall within the average range of the five point Likert scale, being at variance with the observational results which indicated underperformance for these constructs. This could indicate miscalibration of resilience levels (Woods and Wreathall 2008; i.e., learning construct $\alpha = .70$ and SD = .43) within the organization. However, the scores might also be explained by fluctuation in resilient behavior that was observed within and between (senior and junior) rail dispatchers when they were coping with the (dynamic propagation of) random disturbances during real-time operation. A Mann–Whitney U test was conducted to evaluate whether anticipating resilience scores differed between senior and junior rail dispatchers. Although the results indicated that anticipating resilience scores between senior rail dispatchers (Mdn = 11.83) and junior rail dispatchers (Mdn = 10.00) were not significantly different, $U = 30.0$, $p = .608$, $r = .11$, differences in resilient behavior cannot be ruled out completely since common and socially desirable answers could have been given by the rail dispatchers answering the questionnaire. During real-time operation, specific situational demands could elicit differences between junior and senior rail dispatchers based on experience. Situational demands in itself might provide an indication as to why the acceptable degree of internal consistency ($\alpha = .70$; Tabachnick and Fidell 2001) was not met since it could also explain the operational variability observed within individual dispatchers. This essentially reflects the rail dispatchers' notions that no situation is alike, even though situations might appear similar to outsiders since they, for example, both entail disruption handling due to a broken rail switch. In addition to the resilience constructs, domain practitioners assessed the relation-oriented abilities (i.e., shared transformational leadership and cooperation with other teams), which are incorporated into the ADAPTER questionnaire to operationalize the concept of team resilience, as the least well represented within the organization. These results are in line with the resilience observations. Transformational leadership has proved to be a leadership style that effectively stimulates knowledge creation and knowledge sharing at the individual and group levels (Bryant 2003). The fact that that this ability is under-represented (Mdn = 3.14) could affect the learning capabilities of the organization (Zagoršek et al. 2009) and as such explain the performance variability observed. The low rating for cooperation with other teams (Mdn = 2.54) indicates improvement opportunities for handlings across organizational and sub-system boundaries, such as the communication breakdown that occurred in the hooligan example.

4 Discussion and conclusion

Practical implementations and concrete measurement of resilience are a challenging issue that is inadequately addressed in current research practices. The purpose of this research was therefore to take initial steps toward enhancing operationalization and understanding of resilience metrics in the railway sector and quantification of the rail-system resilience state. Overall results indicate that the multi-method approach adopted to establish operational baseline conditions, based on the four system capabilities that comprise resilience, is a reliable method to determine the overall level of rail-system resilience. In addition, WRSs prove to successfully measure deviations from predetermined resilience baseline conditions. Although WRS analysis functions show the potential to enhance understanding of the underlying and complex system dynamics that comprise future resonance, more research is required to determine their full potential.

More specifically, it can be stated that triangulation of the quantitative and qualitative research methods proves to be a useful means of capturing more detail, minimizing the effects of research biases and limitation boundaries of the individual research techniques, and understanding causal mechanisms. The observation (behavioral) and the resilience questionnaire outcomes (attitudinal) both indicated that the system capability 'monitoring' was best represented within the organization and that the relation-oriented abilities that represent team resilience were the least developed, reinforcing the outcomes. Grounding these results within Hollnagel's (2009) framework, a (recurring) pattern emerges that generates insights into the current practices and how this influences the resilience level and operating state of the socio-technical system. The commonality across all real-time dispatching processes is that by means of continuously monitoring and quickly responding, dispatchers try to control the situation and mitigate potential threats. In essence, this reflects exactly the current practice of dispatch activities, monitors the traffic flow and acts accordingly. This generally yields positive and acceptable levels of performance. However, performance variability increases when the 'normal' system functioning (i.e., the corridor team serves as base adaptive capacity; Woods et al. 2014) is challenged and disturbances (external events) cascade across sub-system and organizational boundaries, enhancing the chances for a system decompensation collapse (Branlat and Woods 2010). To prevent such system breakdowns from happening, timely notification and anticipation to incident precursors are crucial. To accomplish this, the theoretical resilience state model for railway systems (Siegel and Schraagen 2014) was implemented to measure deviations from the resilience baseline conditions on the performance system boundary. The measurements in performance variability were translated into WRSs that act as prompts for variables that should be considered. In this process, the observations and resilience questionnaire provide the necessary contextualization that the technical system metrics alone are unable to fully capture. The quantification of WRSs and visualization of cues in a constructive manner help to close the feedback loop and enhance situation awareness. Boosting the relation-oriented abilities within the organization can strengthen these processes and as such reinforce the overall level of system resilience (Hollnagel 2009). Enhancing cooperation and knowledge sharing with other teams could, for example, aid in minimizing information-processing failures (Woolley et al. 2008) where transformational leadership (Bass 1990) could contribute to sense-making, interpretation and understanding of a situation among all members of the (corridor) team (Bartone 2006). In this context, a WRS analysis function should be seen as a means to an end. By providing a means of uncovering potential factors that comprise a WRS, a WRS analysis function can be implemented to guide the WRS selection process and enhance corridor team reflectivity. In this context, reflectivity is defined as the deliberate process of discussing and evaluating team goals, processes and outcomes, learning from failure and successes and craft action intentions for improved future functioning (Ellis et al. 2014; Schippers et al. 2014).

Although results are promising and preliminary feedback from domain practitioners indicates a positive attitude toward implementation of this method, research limitations should be considered. Even though operational parameters were chosen that allow for real-time measurement in the future, the current implementations are based on technological measurement of readily available log data and were created in retrospect. Further empirical research is needed in order to validate and verify these prospects and results during real-time operations. In addition, the fact that we operated in a real-life setting poses a limitation with respect to the replication of the study. Since every rail control post (e.g., the practitioners, the vibe) is different, and the rail control post themselves play a crucial role in the outcome, an exact replication of this study would not necessarily yield similar results. Although it could be argued that this would be possible in a simulated environment, replication in such a literal way was never the main priority for this study. We would rather invite and encourage other researchers to use this study as a base and build upon this work. A potential implication for future work in the line of real-time WRS research would be to create a fully operational advanced graphical user interface design, which can be used to test and capture the complex interactions generated by interrelated components at system level (e.g., usability enhancement based on ecological resilience design principles). Another option is to mature the implementation of WRSs and analysis functions by adding specification criteria and including other resilience boundaries (i.e., safety boundary). Furthermore, it could be examined whether the punctuality WRS and analysis function, which were created to enhance the system capabilities learning and anticipating in the railway system, can be used to enhance these system capabilities in other control room operations of complex socio-technical systems. In addition, it could be interesting to explore these metrics in the broader context of other scientific fields like data science and predictive statistics.

References

Amalberti R (2001) The paradoxes of almost totally safe transportation systems. Saf Sci 37:109–126. doi:10.1016/S0925-7535(00)00045-X

American Association for Public Opinion Research (2015) Standard definitions: final dispositions of case codes and outcome rates for surveys. 8th edition. AAPOR

Back J, Furniss D, Hildebrandt M, Blandford A (2008) Resilience markers for safer systems and organisations. In: Harrison MD, Sujan MA (eds) SAFECOMP. Springer, Berlin, pp 99–112

Bartone PT (2006) Resilience under military operational stress: can leaders influence hardiness? Mil Psychol 18:S131–S148. doi:10.1207/s15327876mp1803s_10

Bass BM (1990) From transactional to transformational leadership: learning to share the vision. Organ Dyn 18:19–31. doi:10.1016/0090-2616(90)90061-S

Belmonte F, Schön W, Heurley L, Capel R (2011) Interdisciplinary safety analysis of complex socio-technological systems based on the functional resonance accident model: an application to railway trafficsupervision. Reliab Eng Syst Saf 96:237–249. doi:10.1016/j.ress.2010.09.006

Berche B, von Ferber C, Holovatch T, Holovatch Y (2009) Resilience of public transport networks against attacks. Eur Phys J B Condens Matter Complex Syst 71:125–137. doi:10.1140/epjb/e2009-00291-3

Bieder C, Bourrier M (eds) (2013) Trapping safety into rules: how desirable or avoidable is proceduralization?. Ashgate Publishing Limited, Farnham

Blandford A, Furniss D (2006) DiCoT: a methodology for applying distributed cognition to the design of teamworking systems. In: Gilroy SW, Harrison MD (eds) DSVIS 2005. Springer, Berlin, pp 26–38

Branlat M, Woods DD (2010) How do systems manage their adaptive capacity to successfully handle disruptions? A resilience engineering perspective. In: AAAI fall symposium, pp 26–34

Bryant SE (2003) The role of transformational and transactional leadership in creating, sharing and exploiting organizational knowledge. J Leadersh Organ Stud 9:33–44. doi:10.1177/107179190300900403

de Carvalho PVR (2011) The use of functional resonance analysis method (FRAM) in a mid-air collision to understand some characteristics of the air traffic management system resilience. Reliab Eng Syst Saf 96:1482–1498. doi:10.1016/j.ress.2011.05.009

Dekker S (2011) Drift into failure—from hunting broken components to understanding complex systems. Ashgate Publishing Limited, Farnham

Dekker S, Hollnagel E, Woods DD, Cook R (2008) Resilience engineering: new directions for measuring and maintaining safety in complex systems. In: Lund University Sch. Aviat. https://msb.se/Upload/Kunskapsbank/Forskningsrapporter/Slutrapporter/2009 Resilience Engineering New directions for measuring and maintaining safety in complex systems.pdf

Doyle JC, Alderson DL, Li L et al (2005) The "robust yet fragile" nature of the internet. Proc Natl Acad Sci USA 102:14497–14502. doi:10.1073/pnas.0501426102

Doyle JC, Francis BA, Tannenbaum AR (2013) Feedback control theory. Macmillan Publishing Company, New York

Ellis S, Carette B, Anseel F, Lievens F (2014) Systematic reflection: implications for learning from failures and successes. Curr Dir Psychol Sci 23:67–72. doi:10.1177/0963721413504106

European Railway Agency (2014) Railway safety performance in the European union. http://www.era.europa.eu/Document-Register/Documents/SPR2014.pdf

Ferreira P, Clarke T, Wilson JR et al (2011) Resilience in rail engineering work. In: Hollnagel E, Paries J, Woods DD, Wreathall J (eds) Resilience in practice. Ashgate, Aldershot, pp 145–156

Furniss D, Back J, Blandford A et al (2011) A resilience markers framework for small teams. Reliab Eng Syst Saf 96:2–10. doi:10.1016/j.ress.2010.06.025

Goverde R, Odijk M (2002) Performance evaluation of network timetables using PETER. In: Allan J, Hill RJ, Brebbia CA, Sciutto G (eds) Computers in railways VIII. WIT press, Southampton, pp 731–740

Hansen IA (ed) (2010) Timetable planning and information quality. WIT Press, Southampton

Heaslip K, Louisell W, Collura J, Urena Serulle N (2010) A sketch level method for assessing transportation network resiliency to natural disasters and man-made events. In: Transportation research board 89th annual meeting, Washington, D.C

Herrmann T, Hoffmann M, Kunau G, Loser K-U (2004) A modelling method for the development of groupware applications as socio-technical systems. Behav Inf Technol 23:119–135. doi:10.1080/01449290310001644840

Hollnagel E (2004) Barriers and accident prevention. Ashgate Publishing Limited, Aldershot

Hollnagel E (2009) The four cornerstones of resilience engineering. In: Nemeth CP, Hollnagel E, Dekker S (eds) Resilience engineering perspectives, vol 2., Preparation and restoration. Ashgate Publishing Limited, Farnham, pp 117–134

Hollnagel E (2011) RAG - the resilience analysis grid. In: Hollnagel E, Paries J, Woods DD, Wreathall J (eds) Resilience engineering in practice: a guidebook. Ashgate Publishing Limited, Farnham, Surrey, pp 275–296

Hollnagel E (2012) FRAM: the functional resonance analysis method: modelling. Ashgate Publishing Limited, Farnham

Hollnagel E (2014) Safety-I and Safety-II: the past and future of safety management. Ashgate Publishing Limited, Farnham

Leveson N (2004) A new accident model for engineering safer systems. Saf Sci 42:237–270

Mach EI (1905) Erkenntnis und Irrthum. Skizzen zur Psychologie der Forschung, Barth

Madni AM, Jackson S (2009) Towards a conceptual framework for resilience engineering. IEEE Syst J 3:181–191. doi:10.1109/JSYST.2009.2017397

Millen DR (2000) Rapid ethnography: time deepening strategies for HCI field research. In: Proceedings of the conference on designing interactive systems processes, practices, methods and techniques, pp 280–288. doi:10.1145/347642.347793

Nulty DD (2008) The adequacy of response rates to online and paper surveys: what can be done? Assess Eval High Educ 33:301–314. doi:10.1080/02602930701293231

Perin C (2005) Shouldering risks: the culture of control in the nuclear power industry. Princeton University Press, Princeton

Pickup L, Wilson JR, Norris BJ et al (2005) The integrated workload scale (IWS): a new self-report tool to assess railway signaller workload. Appl Ergon 36:681–693. doi:10.1016/j.apergo.2005.05.004

Qureshi ZH (2008) A review of accident modelling approaches for complex critical socio-technical systems. In: 12th Australian workshop on safety related programmable systems (SCS'07), Adelaide, pp 47–59

Rasmussen J (1997) Risk management in a dynamic society: a modelling problem. Saf Sci 27:183–213

Scheffer M, Hosper SH, Meijer ML et al (1993) Alternative equilibria in shallow lakes. Trends Ecol Evol 8:275–279. doi:10.1016/0169-5347(93)90254-M

Schippers MC, Edmondson AC, West MA (2014) Team reflexivity as an antidote to team information-processing failures. Small Gr Res 45:731–769. doi:10.1177/1046496414553473

Siegel AW, Schraagen JMC (2014) Measuring workload weak resilience signals at a rail control post. IIE Trans Occup Ergon Hum Factors 2:179–193. doi:10.1080/21577323.2014.958632

Staw BM (1981) The Escalation of Commitment To a Course of Action. Acad Manag Rev 6:577–587. doi:10.5465/AMR.1981.4285694

Tabachnick BG, Fidell LS (2001) Using multivariate statistics. Allyn and Bacon, Boston

Tamvakis P, Xenidis Y (2013) Comparative evaluation of resilience quantification methods for infrastructure systems. Procedia Soc Behav Sci 74:339–348. doi:10.1016/j.sbspro.2013.03.030

Underwood P, Waterson P (2013) Systemic accident analysis: examining the gap between research and practice. Accid Anal Prev 55:154–164. doi:10.1016/j.aap.2013.02.041

van der Beek DFA, Schraagen JMC (2015) ADAPTER: analysing and Developing Adaptability and performance in Teams to enhance resilience. Reliab Eng Syst Saf 141:33–44. doi:10.1016/j.ress.2015.03.019

Weick KE, Sutcliffe KM (2001) Managing the unexpected. Jossey-Bass, San Francisco

Wilson JR, Norris BJ (2005) Rail human factors: past, present and future. Appl Ergon 36:649–660. doi:10.1016/j.apergo.2005.07.001

Wilson JR, Farrington-Darby T, Cox G et al (2007) The railway as a socio-technical system: human factors at the heart of successful rail engineering. Proc Inst Mech Eng Part F J Rail Rapid Transit 221:101–115

Woods DD, Branlat M (2010) Hollnagel's test: being "in control" of highly interdependent multi-layered networked systems. Cogn Technol Work 12:95–101. doi:10.1007/s10111-010-0144-5

Woods DD, Cook R (2006) Incidents-markers of resilience or brittleness. In: Hollnagel E, Woods DD, Leveson N (eds) Resilience engineering. Concepts and precepts. Ashgate Publishing Limited, Aldershot, Hampshire, pp 69–75

Woods DD, Wreathall J (2008) Stress–Strain plots as a basis for assessing system resilience. In: Hollnagel E, Nemeth C, Dekker S (eds) Resilience engineering perspectives, vol 1., Remaining sensitive to the possibility of failure. Ashgate Publishing Limited, Aldershot, pp 145–161

Woods DD, Patterson ES, Cook RI (2007) Behind human error: Taming complexity to improve patient safety. In: Carayon P (ed) Handbook of human factors and ergonomics in health care and patient safety. Lawrence Erlbaum Associates, Mahwah, pp 459–476

Woods DD, Chan YJ, Wreathall J (2014) The stress–strain model of resilience operationalizes the four cornerstones of resilience engineering. In: 5th resilience engineering symposium. http://hdl.handle.net/1811/60454

Woolley AW, Gerbasi ME, Chabris CF et al (2008) Bringing in the experts: how team composition and collaborative planning jointly shape analytic effectiveness. Small Gr Res 39:352–371. doi:10.1177/1046496408317792

Zagoršek H, Dimovski V, Škerlavaj M (2009) Transactional and transformational leadership impacts on organizational learning. J East Eur Manag Stud 14:144–165