Human performance in rail: current assessment and future opportunities

Abigail Fowler[1], Dr David Golightly[2], Prof Sarah Sharples[3], Dr Catherine Harvey[4], Dr Max L. Wilson[5], Dr Huw Gibson[6]

[1,3,4,5]University of Nottingham, [2]Newcastle University, [6]RSSB

ABSTRACT

Automated and assistive technologies are being introduced to the rail industry to support the tasks of drivers and signallers. As rail tasks become less physical and more cognitive, measurement of human performance and evaluation of the impact of these new technologies needs to evolve. As pressure to achieve high performance increases, the importance of understanding human performance limits also increases. This paper on human performance in rail is based on semi-structured interviews with operational and human factors specialists. The paper builds a 'where are we now' picture of current staff performance assessment in rail. It considers in addition 'where could we go next?' using developments in measurement technologies, such as physiological measures, and research in both Air Traffic Control (ATC) and automotive industries. A thematic analysis of interviews identified three key themes: people, time and attributes. Firstly, the performance of a wide range of people can impact rail operations, which highlights how complex a socio-technical system rail is. Secondly, staff assessment occurs at certain times in rail (recruitment selection, routine competency assessment and following an incident). Thirdly, beyond task competency, individual attributes can indicate good or poor human performance. Future opportunities include the increased use of logged data, physiological measures, and simulators, to improve our understanding of human performance in rail. An important consideration is how the data will be used and combined with wider competence and performance assessment processes.

KEYWORDS

Human Performance, Competency, Rail Industry

Introduction

Demand for rail travel has been increasing, with passenger journeys in 2017-2018 up 28% compared to 2007-2008 (Office of Rail and Road (ORR) 2018). To meet this increase in demand, the rail industry is increasing the capacity of the rail network by allowing more trains to use the tracks at any one time. To achieve this, automated and assistive technologies are being introduced to support staff performing safety critical tasks. These staff have responsibility for the safety of themselves, colleagues, passengers and the public. Their tasks are increasingly cognitive rather than physical, as more automation is introduced, making the effort required more difficult to assess through physical measures or observation (Sharples, Millen, Golightly and Balfe 2011). The challenge is how to measure human performance and assess the impact of these new technologies.

The term 'human performance' refers here to the act of performing and achieving a task, as distinct from performance shaping factors such as fatigue, task load, health or wellbeing, that can impact the task performance. This paper focuses on three research questions: how is human performance currently measured in rail? how is human performance measured in other transport industries?, and how could human performance be measured in future to assess the impact of new technologies? The paper initially considers the broad range of people who impact rail operations, then concentrates on drivers and signallers. These roles were chosen as examples of front-line staff making time critical safety decisions, whose tasks are increasingly cognitive, less physical and more sedentary.

Automotive and Air Traffic Control (ATC) industries are considered here to compare how human performance can be monitored and assessed in driving (car and train) and control tasks (air traffic controller and signaller). These industries were chosen for their similarly increasing automation and more cognitive, sedentary tasks. Advancements in technology provide an opportunity to assess these changing roles in new ways. Current self-assessment workload measures require interruption of a task or application only after completion of a task (Sharples and Megaw, 2015). Physiological measures offer the potential for continuous data without interrupting the task. Physiological measures detect aspects of physical activity, such as heart rate variability, or Galvanic Skin Response (GSR), that can be used to infer levels of cognitive activity. This paper considers whether such technologies could be applied in future to assess human performance in rail. An important consideration is how the data from these technologies will be used and whether they can fit with wider competence and performance assessment processes.

Method

This study applied a qualitative approach to discover how staff performance is currently assessed in rail. The intention was to gain a wide range of experience and opinions from specialists across rail and from comparable industries. Semi-structured interviews were conducted with fourteen participants from the following domains: twelve rail; 1 automotive; 1 ATC. Rail specialists represented a range of operational job roles and experience including Head of Drivers, Head of Operations, Operations Standards Manager, Train Driver, Signaller, Human Factors Specialist, Accident Investigator, Systems Modelling and Simulator Expert. Some individuals had multiple areas of expertise. Representatives from both Train Operating Companies (TOC) and a Freight Operating Company (FOC) were included. In addition to the interviews, a TOC meeting of operational managers was attended which covered driver training, recent incidents and current operational issues. During the interviews broad initial questions were asked including: what current challenges in rail rely on the performance of humans (staff or other); what data are currently collected in rail that capture human performance of drivers and signallers; how are data collected; and who uses the data? Follow on questions were asked to encourage specialists to explain their answers and give examples.

Interviews were conducted between March and October 2018. Eleven interviews were conducted face to face, four over the telephone. All interviews were audio recorded, with notes taken during the interview to guide questions and capture themes. Interviews were planned to last 45 minutes, the actual interviews ranged between 29 – 87 minutes and averaged 54 minutes. Emerging thematic analysis (Strauss and Corbin 1990) was used to identify categories and broad themes. This was completed by a detailed analysis of the first five interviews to build a coding template. These were transcribed, printed and coded initially on paper using coloured pens, then iteratively in NViVo. The resulting themes were then applied to the remaining 9 interviews.

Results

The thematic analysis of interviews identified three key themes relating to human performance in rail: People – who impact rail operations; Time – when competency is assessed; and Attributes – Indicators of good or poor performance. These will be presented, followed by a summary of current research in rail, ATC and automotive industries identified in interviews, followed by what future opportunities these offer for the assessment of human performance in rail.

Before presenting these findings, it should be noted that in interviews it became apparent that the term 'performance' refers in rail to operational performance and 'human performance' is understood more as staff 'competency'. In this paper performance is referred to as operational or human performance, and 'human performance' is used as a broad term that includes competency.

People who impact rail operations

The performance of a broad range of people can impact rail operations. One specialist, when asked where in rail humans impact rail operations, answered "everywhere" (P11). The following people were mentioned in interviews, with categorisation added here based on their visibility to passengers and the industry definition of Safety Critical Tasks (ORR 2017):

- Front-line staff most visible to passengers: Driver; Guard; Train Dispatcher, Track workers
- Staff performing Safety Critical Tasks less visible to passengers: Signaller; Level Crossing Operator (including CCTV); Electrical Controller; Mobile Operations Manager (MOM); Maintenance Fitter; Shunter; Axel Inspector
- Managers: Managers of front-line staff; Standards Manager; Train Services Manager; Operations Manager; Head of Engineering; Head of Safety; Station Controller (who allocates platforms); Shift Signaller Manager; Track Section Supervisor; Train Running Controller; Route Control Manager; Route Control Incident Manager
- Other staff: Timetable Planners; Control Centre Technicians (receive fault alerts from remote condition monitoring)
- Suppliers to operations: Signal Designer; Traffic Management System Designer; manufacturers; funders; government; Rolling Stock Leasing Companies (ROSCOs).

Whilst this paper focuses on the impact of staff, other people play a role:

"Road users at level crossings have a role to play. Road users at level crossings and the platform train interface are two bits of the railway where the performance of members of the public are important, and these are people we do not directly control." (P2)

This wide range of people shows how complex a socio-technical system rail is. In terms of assessment, drivers and signallers have high levels of assessment and are two of the biggest groups of staff. This paper will therefore focus on drivers and signallers. It should be noted however that their ability to achieve good performance is influenced by the performance of others including passengers and the general public outside of their control.

Time – when competency is assessed

Comments revealed concerns about driver and signaller competence at different career stages. Assessment predominantly occurs at selection during recruitment, routine competency review; and following an incident. Each will be presented in turn.

Selection

Selection during recruitment includes an interview, medical, and psychometric tests (drivers only). One representative TOC (P9 & TOC meeting) felt that the standard of candidate had reduced in recent years. More applicants now pass their assessment, but then fail basic training. The TOC suggested this increase had been since a change to the industry's standard Initial Driver Assessment in 2013. The TOC had four drivers fail training on one course of 8 candidates (P9, TOC meeting):

"We've had more failures through training in the past few years than we've ever had and at least 25% of our incidents now are down to post qualified drivers" (P9)

There is also anecdotal evidence of a change in applicant demographic:

"Youngest freight drivers are late twenties, mid-thirties, early forties. Probably as a second career sometimes (having) been in the army, or police, or teachers. We've had a bank manager who's joined because it's good money… So I think the demographics are going to change over the years." (P13)

"We lost people who had been in the industry for 20-30 years and spent the weekend looking at trains at the end of platforms. Then (new) junior people are doing it as a computer game in a sense, who have to learn the real world. It takes a long time for them to get to speed." (P5)

One specialist suggested the pay offer could be encouraging this change in applicant (P7) and implied that this did not improve the suitability of individuals for the role. These findings may indicate what motivates individuals to apply to rail is changing, resulting in people with a wider range of previous experience and interests working in rail. Some individuals may have more intrinsic reasons for applying such as an interest in rail, others may be motivated by more extrinsic factors such as pay and conditions.

Competency Assessment

Once qualified, staff competency assessments occur on a rolling programme over 1 - 2 years (P7, P9). Signallers are observed at their workstation and drivers during a cab ride. Samples of voice recording are checked for correct use of safety critical communications (P8). Driver monitoring includes samples of data from the On Train Data Recorder (OTDR) to check performance such as speeding (P7) and provide feedback on good performance (P9) e.g. correct sequence of actions (P7). Currently signaller workstation control logs are not routinely reviewed (P6).

Two specialists (P3, P13) questioned whether staff perform differently when observed. To address this a FOC has moved away from relying on observations:

"The primary method of assessing a driver is a download (from) data recorders, because they (a driver) will drive nearly perfectly if someone's sat next to them… your data recorder assessment is probably warts and all. The detail you can get to is quite astonishing." (P13)

A FOC's safety team receive live OTDR alerts e.g. if a driver leaves a work site without turning the Train Protection Warning System (TPWS) back on before driving over 40 mph. Patterns in OTDR data can also show erratic actions such as use of the Drivers' Safety Device (DSD) (P13).

Simulators are used for training, such as drivers' route learning (P9) and to practice degraded working (P13), however individual performance is not assessed during these sessions (P7).

Incident Investigation

Staff are interviewed following an incident e.g. a Stop Short Door Release (when only part of a train is at a platform and passengers risk falling onto the track). Various sources of data can support further investigation. Investigators can check staff received everything they needed (P8) e.g. training. OTDR data can identify driver errors or confirm their account of events. The quality and ease of access to OTDR varies across the industry. Signallers' inputs are logged, but only at modern electrical control centre workstations (P6, P8). These logs do not identify the individual so linking data to an individual requires checking their roster.

Factors that influence performance are also considered such as fatigue and levels of experience. Participants reported drivers rarely report feeling fatigued (TOC meeting), although it is known to be an industry issue (P2, P13) that contributes to incidents (P13, TOC meeting). In terms of experience, one TOC estimates that 25% of their incidents are by drivers in their first year (P9).

Participants felt there is an element of chance as to whether an individual has an incident (P2) and noted individual performance has the potential to avoid an incident as well as contribute to one.

Attributes – indicators of good or poor performance

Responses in interviews focused around competency and incident investigation. What made someone good at a role seemed difficult to define (P4). Participants were therefore asked what made a good driver (as an example role). Their responses are presented in Table 1. The skills match requirements identified by the Risk Based Training Needs Analysis tool (RSSB 2018).

Table 1: Attributes of good and poor driver performance

Good driver attributes	Poor driver attributes
• Turns up • Concentration • Methodical approach • Preparation & Planning • Situational Awareness • Excellent Non-Technical Skills • Uses Risk Triggered Commentary Driving (point and call out loud) • Concentrates from station to station, taking each stage at a time • Can take on lots of information • Can work solo, are self-contained, introvert • Intelligent without getting distracted or their mind wandering • Practical skills (e.g. train inspection) • Fewer incidents/ near misses but may not be zero incidents • Every day is a school day attitude	• Complacency • Fails to use Risk Triggered Commentary Driving • Finds it difficult to take on lots of information • Extrovert • Life style - they cannot deal with long periods of work • Greater number of incidents/near misses • 'What's in it for me' attitude (but this does not mean they have incidents) • A 'square peg in a round hole'

Whilst physical attributes, gender and age were not mentioned, application of skills, personality and attitude were mentioned. Demographics were viewed as quite superficial (P2). It was noted that experienced staff can find the transition to newer technologies, such as tablets, difficult (P13), with some choosing to retire rather than transition. A level of uncertainty was detected in interviews as to whether staff with experience of older ways of working, or newer recruits who only learn the newer technology, are better able to achieve good performance. One participant described younger signaller applicants as 'gamers' who have a different mindset, with the benefit of adapting quickly and positively to digital technologies (P14).

The number of incidents an individual has was not deemed a reliable measure (TOC meeting). It is a point of debate for accident investigators how much previous incidents are relevant to a current investigation (P11). TOC operational managers expressed confidence in some drivers who had had an incident and a lack of confidence in some drivers who had not had an incident.

In terms of attitude of staff:

"Nobody goes out there to have an accident, nobody goes out there saying "I'm going to mess up the system today". They're all just trying to get the job done and go home at the end of the day" (P11)

It was noted that demands for precision are increasing (P1) and staff are close to the limit (P5):

"The requirement for a train driver now is that you need to be 100% accurate 100% of the time and there isn't any let up to that. Twenty five years ago there wasn't the pressures in the performance, capacity, and the safety requirements" (P1)

Interviews highlighted challenges beyond individual roles. Firstly, due to how long technology deployment takes, staff must constantly operate with a mixed age of technologies. Secondly, across the industry corporate knowledge is lost when staff leave the industry (P5, P14) leading to decisions being made that impact front line staff by individuals with limited understanding of the implications. Thirdly, some roles have become more mundane (P13): driver routes are not varied (P7), trains have become easier to drive, with more comfortable cabs (P1) and automation protects staff actions (P4). These all have the unforeseen risk of leading to inattention due to underload (P2), (TOC operational meeting). One participant questioned how individuals can maintain focus (P7).

Current Research in Rail, Air Traffic Control and Automotive Industries

A different point of view was noted between operational and human factors specialists in rail. Operational staff talked about competency management. Human Factors specialists mentioned research involving other ways of assessing performance including: a tool for determining signalling task demand (Pickup, Wilson & Lowe 2010); OTDR data showing different personal driving styles (P8); driver GSR correlating with difficult conditions such as risk of train slipping in low adhesion conditions (P8) (Crowley & Balfe 2018); and a tool to assess exposure to risk such as the Red Aspect Approach Tool (RAAT) (P2) to indicate how frequently drivers are exposed to the risk of a Signal Passed At Danger (SPAD) (P1).

In ATC subjective measures are routinely used including workload, confidence and Situational Awareness to assure the system is safe (P12). Objective data comes from mouse clicks, and time on radio can indicate complexity of task. Data identifies the sector, not the individual, and are entered quickly by staff on an iPad at the end of a session. In addition, controller's biometric data are collected including their visual scanning patterns (using eye tracking) and how they deal with stress (using a chest strap to detect heart rate variability, and GSR from a device worn on the arm). These measures are used in simulations, training and live environment. Intended changes to the system can therefore be assessed through design and deployment stages. In the live environment the data from physiological devices can provide the supervisor real time indicators and a decision support tool for when the task load of a controller needs adjusting. In ATC task load can be adjusted by rerouting aircraft to different sectors. In addition to the benefits to maintaining operations, these measures monitor individual consistency and spot change. The participant gave the example of an individual whose Electroencephalogram (EEG) reading was different from normal, and the next day they were ill. The change in EEG showed a physiological change before the individual was aware of becoming ill.

The automotive industry is developing real-time driver monitoring features (P10). Products such as fatigue monitoring are available at the high end of the market e.g. a dash board coffee cup indicator to recommend a driver takes a break when they show signs of fatigue (P10, Car Sales 2018). Cameras are being introduced to detect aspects of the face including eye position to monitor slouching (P10). Using emotional expressions is still in development as they vary between individuals. Cars are available that can also monitor driving performance such as lane keeping (referred to as lateral driving performance). Features that are deemed to be for safety are generally well received by customers, although attitudes to being monitored vary between cultures and between generations (P10). Fatigue mitigation techniques is another area being researched. A study of long-haul lorry drivers noted that drivers made phone calls to help combat underload (P10).

Future Opportunities

Participants suggested ways human performance could be assessed in future. Firstly, using existing data in rail to calculate 'exposure to risk' rates e.g. take timetable information of how many times a driver stops at a station (P1) to determine exposure rate to incidents of Stop Short or Failure to Call. Secondly simulators in rail are currently used for training (P8) but not widely for assessment. Currently only about 5% of the data collected from a simulator session are used (P14). Simulators could be used to study aspects of performance such as reaction time. OTDR could also be used (P3) to record an 'optimal' drive by an experienced tutor to show to trainee drivers (P13). Finally, wearable technologies could be explored further to determine which physiological measures could indicate human performance in rail.

Discussion

Current assessment of human performance in rail is around competency and occurs at certain times. Data focuses on human equipment inputs (e.g. OTDR data), less on individual performance. One challenge in rail is that as roles change, appropriate competency is changing without an associated change in assessment. These findings and challenges are summarised in Figure 1.

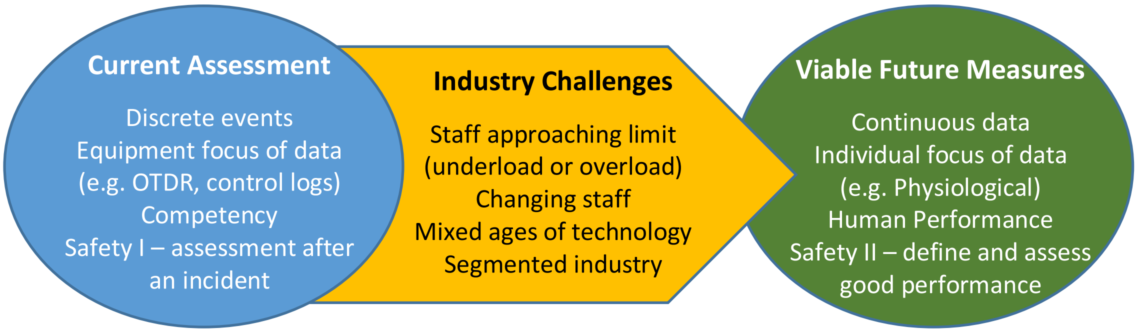

Figure 1: Challenges and future opportunities for human performance assessment in rail

If incidents occur a more detailed data analysis is completed. This fits with the Safety I approach (Hollnagel 2014), which tends to be reactive and can lack the awareness to predict incidents. As one participant said rail has "no way of knowing how close to the limits of ability staff are" (P5). Adopting a Safety II approach in future would focus on identifying what factors contribute to good human performance and result in successful operational performance. These factors include the individual attributes of skills and knowledge, as well as how an individual applies them.

The terminology used when discussing performance needs further thought. Whilst the phrase 'human performance' is not used in rail by operational staff, or seen as a distinct issue, participants did recognise the performance of a wide range of people impacts rail, including passengers and public. There were also examples of good work and research around human performance, but a lack of awareness of this across the whole industry. This fits with the finding the industry is segmented, with priorities varying across operations, locations and time scales.

In considering future opportunities, increased use of data logs (e.g. OTDR) and simulators are possibilities in rail. ATC, rail and automotive industries are exploring the potential of various physiological measures. These measures offer a potential future opportunity to collect and analyse continuous data in rail, without interrupting the task, and measure good performance rather than wait for an incident to occur. It will be important to assess the suitability of these measures, practicalities, and ethics of collecting data that measure individual performance of staff. Further research should include consideration of the attitudes of staff to the use of such data, who would have sight of it, and how such data could be used within wider competence and performance assessment processes.

References

Car Sales (2018) Fatigue monitoring systems explained. 5 March 2018. https://www.carsales.com.au/editorial/details/fatigue-monitoring-systems-111391/

Crowley, K. and Balfe, N., (2018). Investigation of Train Driver Physiological Responses. 32nd Human Computer Interaction Conference. hci2018.bcs.org

Hollnagel, E., 2014. *Safety-I and Safety–II: The Past and Future of Safety Management*. Ashgate Publishing, Ltd.

ORR (2007). Safety Critical Tasks – Clarification of ROGS Regulations Requirements. Railway Safety Publication 4

ORR (2018) Passenger Rail Usage 2018-2019 Q1 Statistical Release http://orr.gov.uk/__data/assets/pdf_file/0003/39342/passenger-rail-usage-2018-19-q1.pdf [Accessed 23.10.18]

Pickup, L., Wilson, J. and Lowe, E., (2010). The Operational Demand Evaluation Checklist (ODEC) of workload for railway signalling. *Applied ergonomics*, *41*(3), pp.393-402.

RSSB (2018) Risk-based Training Needs Analysis (RBTNA) www.rssb.co.uk [Accessed 30.

Sharples, S. and Megaw, T., (2015) Definition and Measurement of Human Workload. In Sharples, S. and Wilson, J.R., 2015. Evaluation of human work. pp515-548.

Sharples, S., Millen, L., Golightly, D., & Balfe, N. (2011). The impact of automation in rail signalling operations. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 225(2),179-191.

Strauss, A. and Corbin, J.M., 1990. *Basics of qualitative research: Grounded theory procedures and techniques*. Sage Publications, Inc.

The author is supported by the Horizon Centre for Doctoral Training at the University of Nottingham (UKRI Grant No. EP/L015463/1) and by a rail consortium external partner including Railway Safety and Standards Board (RSSB), Network Rail, and Rail Delivery Group (RDG).