Attitudes to physiological wearables in the workplace in the railway industry

Abigail Fowler[1], Catherine Harvey[1], Max L Wilson[1] & Sarah Sharples[1]

[1] University of Nottingham, UK

SUMMARY

This study evaluates railway staff attitudes and perspectives to physiological wearables in the workplace. Findings indicate wearables could suit use in live operational environments, provided data use and data privacy concerns are addressed. The successful application of wearable measures relies on both data being relevant, and wearables being acceptable to staff. The study focuses on signallers, with implications for other staff in transport industries.

KEYWORDS

Physiological wearable measures, Railway signallers, Staff attitudes

Introduction

Railway signaller Mental Workload (MWL) varies within and across long shifts. Measuring MWL in live railway operations is constrained as observations are of limited duration and self-reported ratings risk distracting staff. This research explored whether physiological wearables could measure MWL in live railway signalling operations. MWL is a concept that encapsulates task demand, individuals' experiences of workload, and task performance (Sharples 2019). Previous research suggests MWL could be inferred from changing physiological state (Charles & Nixon 2019). Despite the growth in wearable technologies, little research has been conducted on staff acceptability (Gribel, Regier and Stengel, 2016). Staff attitudes and perspectives were explored to understand the feasibility of applying wearables.

Electrodermal Activity (EDA) and Heart Rate Variability (HRV) were selected as previous research identified the potential for them to indicate aspects of MWL. EDA reflects activation of the sympathetic nervous system, whilst HRV reflects both the sympathetic and parasympathetic nervous system activity (Tortora and Derrickson, 2007). Increased EDA implies stress and increased alertness (arousal) (Healey and Picard, 2005). Peaks in EDA, known as Skin Conductance Responses (SCRs), have implied anticipation in train drivers (Crowley and Balfe 2018) and varied depending on the content and implication of signaller phone calls (Broekhoven, 2016). HRV measures the varying time gap between R wave peaks in electrical heart activity. Low HRV correlates with high workload (Lehrer et al., 2010) and train drivers' HRV reduced at stops and tunnels (Song et al., 2014). To determine the potential to collect this data in a live operational environment, this study investigated which measures could suit live operations and which factors contribute to staff attitudes to the use of their data at work. The study investigated attitudes towards three measures including the devices and the data they collect: a wearable wrist strap detecting EDA; a chest strap to detect HRV; and an app on a mobile device to collect self-report workload using the existing Integrated Workload Scale (IWS) (Pickup et al., 2005). To understand staff attitudes to these different measures the study considered perceived ease of use and perceived usefulness (following the Technology Acceptance Model extension (TAM2), Venkatesh & Davis 2000) and other factors including comfort, distraction, and trust (Gribel et al.; Jacobs et al., 2019).

Method

Semi-structured telephone interviews were completed with 18 signallers aged 28 – 67 years (M = 46.9, SD = 12.9) with an average of 14.2 years' experience (SD = 11.1). The average length of interview was 60 minutes (SD = 14). Participants were signallers and shift managers at East Midlands Control Centre, including a local Union representative. Participants were recruited using a snow-ball sampling method via poster and word of mouth. Participation was voluntary with no incentive offered. The study received ethics approval from the University of Nottingham and representative railway Unions were informed prior to the interviews. Participants completed a consent form and demographic questionnaire online.

Table 1: Interview topics and post-interview statements. G = Gribel et al., J = Jacobs et al., P = Parasuraman and Colby, U = Urquhart and Craigon, V = Venkatesh and Davis, and W = Wolf et al.

Interview Topic	**Adapted from**
Experience – own experience of wearables	J, V, G
Experience – workload examples	New item
Perceived Ease of Use – Device distraction	P, G, J, U
Perceived Ease of Use – Device comfort	U, W
Subjective Norm/Image – Likely reaction of colleagues	U, V
Anonymity of data	J, U, W
Trust – concerns about data use	J, P, G
Data collection time needed versus tolerable	New item
Perceived Usefulness - How useful would it be to…	
Use Case A: demonstrate to others how hard you work?	New item
Use Case B: understand your data?	New item
Use Case C: demonstrate to others impact of changes?	New item
Job Relevance - What's inferred – which most relevant to …	
Use Case A: assessing task?	V
Use Case B: assessing trainees?	New item
Use Case C: assessing impact of change?	New item
Just because we could use these measures, should we?	U
Post-Interview Statements and Scale (1 strongly disagree – 7 strongly agree)	
Measuring individual signaller workload is important in rail	V
Measuring my workload is relevant to my job	V
Wearing devices wouldn't require a lot of my mental effort	V
A lot of my mental effort would be required to interact with the devices	V
I would find the devices difficult to use	V
During a shift the devices would be distracting	G, J, P
The devices could be a status symbol in my organisation	V
I wouldn't use the devices because I would be concerned about being tracked	W
Assuming I have access to the devices, I intend to use them	V
Given that I would have access to the devices, I predict I would use them	V
I would not recommend the devices to my colleagues	New item

Interview topics and the post-interview survey were adapted from technology acceptance and wearable technology research (Table 1): Technology Acceptance Model extension (TAM2) with perceived ease of use and perceived usefulness including whether it is a status symbol (Venkatesh & Davis); TAM adapted to wearables including concern over being tracked (Wolf, Menzel, and Rennhak, 2018); Technology Readiness Index (Parasuraman and Colby, 2015); Moral IT deck including privacy, ethics, legal and security factors (Urquhart and Craigon, 2021); acceptance of wearables computing including sharing data with third parties (Gribel et al.); and staff acceptance of wearables (Jacobs et al.).

Prior to interview, participants received an 'Introduction to Wearables and App' information sheet about the three measures: a wrist strap to detect EDA (Affectiva Q^{TM}); a chest strap to detect HRV (Zephyr Bioharness) and an app on a mobile device to collect self-report workload using the IWS (Pickup et al.). In interviews use cases prompted discussion on hypothetical future uses that current staff could relate to (Table 2). These were developed from industry interviews, a scoping review of physiological data and simulation study (Fowler, 2023), and technology acceptance research. The use cases included: Perceived Usefulness (Venkatesh and Davis), presenting a range of potential uses; Job Relevance, with most relevant aspects of workload and physiological data; Anonymity and Trust (Jacobs et al., Urquhart and Craigon), including third-party access (Gribel et al.); and Time Required, which considered feasible durations, based on consultation with industry. A post-interview survey asked participants to rate their agreement with a set of statements on a scale.

Table 2: Use case prompts used in interviews

	Use Case A Understand Signaller Workload	**Use Case B Learning Aid**	**Use Case C Assess impact of new technology or procedure**
Perceived Usefulness	Detect peaks and troughs in workload and effort	Track progress, self-learning, assess training	Assess effectiveness of change
Job Relevance	Infer anticipation, alertness, stress, time pressure, brief peaks	Infer alertness, confidence, unexpected event, stress, effort	Infer stress, effort, unexpected system responses
Anonymity & Trust	Anonymised	Trainee data shared with trainer and device supplier	Labelled with initials, shared with manager and investigator
Time Required	Data collected over 1-2 shifts	Data collected before/during/after training	Data collected before/during/after change

Interviews were conducted over the telephone, transcribed, and coded to identify factors underlying staff attitudes (Saldaña, 2016). The coding and analysis combined an inductive and deductive approach to reflect the original deductive prompts in the interview topics and use cases, and themes that emerged during the interview process and coding. Half the interviews were coded on paper producing a framework of 54 codes and categories. A second analyst completed a review by sorting codes into categories without reference to the framework. All codes were retained, some categories were merged, and one new category 'Data Quality Uncertain' created. This new coding framework was applied all interviews in NVivo 12 (QSR 2019).

Results

The coding process produced three themes: Justification; Data Collection; and Consequences (see Table 3). These include a range of perspectives and attitudes, from endorsement to opposition.

Theme 1: Justification

This theme includes the benefits of measures, relevance of data, and social acceptability. All relate to how well wearable measures currently fit into the railway industry.

Wearable measures were seen as having potential, if they provide tangible benefits to railway operations such as leading to improvements in effectiveness or safety. Another benefit was improving the visibility of staff effort, as the job can require significant mental effort whilst looking easy to observers; *“I think it [MWL measurement] is very useful in certain senses. I think people in different areas of rail operations, we don't see what they do, and what their stress levels are, and they don't see what the signallers’ side is.”* [P8].

Table 3: Coding framework for attitudes

Themes		
1 Justification	**2 Data Collection**	**3 Consequences**
Tangible benefit to railway operations	Comfortable devices suit live operations	Importance of trust in those who see their data
Data relevance to signalling task	Data quality uncertain	Suspicion data privacy will not be kept
Low social acceptability	Risk of distraction	Concerns over data misuse

Participants overall held positive views around the use of wearables to highlight variance in workload across a shift, the impact of incidents on staff, and how trainees react to certain situations compared to experienced signallers. Furthermore, there was a positive perception that data from wearables could inform tailored debrief sessions for staff to become more conscious of what they are doing. Potentially both trainer and trainee could benefit from wearing a device, generating comparative data that could be used in post-task reflection.

Regarding data relevance both HRV and EDA data were deemed relevant to signaller workload if they could imply stress, confidence, and alertness. Stress can occur with failures of equipment or incidents, such as a points failure, children on the track, or trains disappearing from screens due to leaf fall covering the tracks. Drivers only call a signaller if there is a problem, so hearing the phone ring can be stressful: *“All of a sudden you’ve got an emergency call to deal with and you’re finding out whether or not a driver/ a train has actually run over someone”* [P1]. Not all unexpected events are stressful however, so any brief peaks in MWL would not affect signallers for long.

Participants reported feeling confident when everything runs smoothly and on time. Trainee confidence increases, and stress decreases, as they build experience of successfully dealing with events; *“Confidence. I believe it’s a big thing in this job. And knowledge is confidence. You know if you've got the knowledge, you ARE more confident. And that gives you a much better set of skills to work the workstation.”* [P15]. A difference in alertness was reported between trainees and experienced signallers; *“Because you’re new in the role and you're there like a meerkat, because you’re constantly looking, you’re always then anticipating what's going to happen next? [Compared to] an elephant or rhinoceros or something like that that's possibly not as threatened by predators, something that's more laidback. So the alertness level is there because they ARE looking for the dangers around them, BUT they’re not up on their hind legs scouring this, that, and the other on a workstation”* [P16]. Findings suggest confidence builds and alertness peaks reduce over time with experience.

Time pressure and anticipation were deemed less relevant to signallers. As staff can manage their own workload (by holding trains at signals), time pressure was less relevant. Anticipation was viewed as bad for a signaller if they made assumptions. Instead, they advocated a level of preparedness and responding promptly; “*You can NEVER expect to have a problem. It's just knowing how to deal with that problem at that time.”* [P16].

Staff were asked to report, using rating scales, their predicted attitudes to wearable devices recording physiological data, and an app on a mobile device to record self-reported MWL. Most staff predicted they would use one or more device (see Figure 1). Low social acceptability was predicted amongst staff however, with 83% of participants anticipating some level of resistance to the new measures. Wearables were not viewed as a status symbol, with one signaller commenting there could be “a little bit of banter” from colleagues for wearing one [P3].

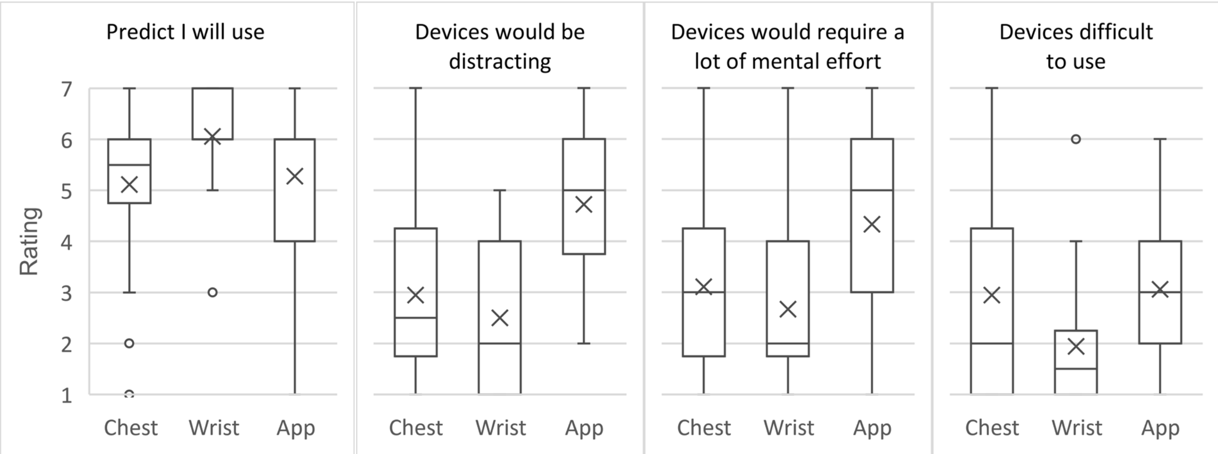

Figure 1: Staff questionnaire responses to devices in box charts. Ratings are from 1 strongly disagree to 7 strongly agree. Whisker is 1.5 times interquartile range, and dots are outliers.

Theme 2: Data Collection

This theme includes comfort, data quality and distraction relating to the devices and data collection.

The qualitative data indicates that the two wearables would be more suitable in live operations than the app, as the app was rated the most distracting and require the greater mental effort (see Figure 1). The wrist strap was rated high comfort, low distraction ratings and least difficult to use (see Figure 1). Signallers commented the chest strap was a bit strange, would require “more messing about putting it on”, and may be an irritant or uncomfortable, but it was difficult to predict comfort; *“I don't know, (laughs) but obviously without trying without trying the chest one it's hard to say”* [P4]. The chest strap may suit use after a period of familiarisation. Whilst signallers may initially be conscious that they were wearing them, this would be likely to reduce once they were used to them.

Staff were uncertain about data quality. Positives included the perception that wearables would provide more objective data than self-reported MWL with fewer false or missed ratings, particularly when busy or for a while after an incident. Staff queried how accurate physiological data would be for workload. Participants understood that stress increases heart rate but queried whether it was clear what was good or bad levels in terms of workload.

Regarding distraction, participants reported the wrist strap as the least distracting and the chest strap may be intrusive if staff are conscious of wearing it. The app was the mostly distracting, especially for trainees, and this matched survey rankings (Figure 1); “*On a live workstation, we've got safety critical situations, we're asking them to possibly, just by doing this (answer the app) to distract them from what they're already doing”* [P14]. Staff may be distracted wondering what the physiological data was showing, particularly during an incident. Despite this, ratings were low for concern about being tracked in the post-interview survey.

Theme 3: Consequences

This theme includes trust, data privacy and data misuse, and relates to how the data is used.

The importance of trust in those who see signallers' data emerged as a key factor. Signallers were cautious about sharing within their organisation. Only 11% of responses were positive regarding sharing named data with managers. One reason given was staff were concerned managers could misunderstand or misinterpret their physiological data; *"If they ((managers)) don't fully understand what's going on, then I'd rather you ((a researcher)) come and tell me"* [P1]. Staff seemed more prepared to share confidential data with researchers or anonymised data with third party suppliers.

Seven participants raised data privacy concerns including word spreading; *"You don't want to walk into work and then everyone saying 'Ohh we've seen your stress levels yesterday'"* [P6]. Staff would prefer their data to be anonymised. Whilst data could be shared confidentially between researcher and individuals, staff would not want to be visible to staff and managers. In the work setting maintaining data privacy is difficult however, as staff identities may be implied from anonymised data from the workstation (such as location and event) with named data in the roster. Maintaining anonymity is also a challenge when data is collected from only a few staff or trainees.

Concerns over data misuse were raised by 14 participants. A significant concern was data misuse may lead to loss of employment; *"Oh the company will get rid of me then if they don't think I could do the job properly"* [P11]. They thought colleagues would not want this added pressure. Participants were concerned managers would criticise staff about their physiological data, as has happened with delay attribution or wrong routes. Participants did not want the data used for disciplinary matters, to assess job performance, or as a negative indicator of their capacity. Another concern was staff may be unfairly compared to each other. Differing levels of stress or heart rates between signallers were not perceived as relevant if the performance outcome was successful. In cases where data is not used to incriminate staff in any way, and kept only for research, then signallers were more positive: *"I'm all for it and I'm quite happy with the technology, as long as they are USED by the company in the correct way. That's vital, that's vital."* [P14].

Discussion

Attitudes were most positive to the wrist strap device in terms of comfort and (least) distraction. The app raised the greatest concerns over distraction, particularly when used by trainees. These findings match previous research on comfort (Urquhart and Craigon; Wolf et al.) and distraction (Parasuraman and Colby; Gribel et al.; Jacobs et al.; Urquhart and Craigon). Measuring workload and inferring stress, confidence, and alertness were identified as most relevant to signallers. This indicates both wearable measures have potential to be useful in the railway industry.

Factors identified in theories that were not found to be the case in this study included measures not being viewed as a status symbol (Venkatesh and Davis), less concern over being tracked (Wolf et al.) and being accepting of sharing data with third parties (Gribel et al.) including researchers and device suppliers. Participants' concerns over data misuse could explain why they did not view measures as a status symbol. Staff being used to existing high levels of monitoring, such as phone calls, could explain lower concerns about being tracked being reported in the survey. Participants' higher trust of those outside the organisation, specifically researchers and device suppliers, may reflect the reduced influence over staff employment.

Trust in data use emerged as a key factor matching theories in literature (Parasuraman and Colby; Gribel et al.; Jacobs et al.). No devices would be worn by staff if they were concerned managers would use the data to criticise, blame, or assess their job performance. A significant concern was the misuse of data could lead to loss of employment. The findings suggest that currently an implicit distrust exists in sharing data. This concern is removed if data is anonymised, which matches existing theories (Jacobs et al.; Urquhart and Craigon). Maintaining data privacy is difficult in practice however, as an individual's identity at work may be inferred from their anonymised data and the published roster. Location information and type of incident could be sufficient for colleagues or managers to determine who was on shift. Also, to use physiological data as a MWL measure requires individual and task events information to understand individuals' MWL experience. Then to inform positive individual support such as flexible breaks or tailored training, rather than remaining confidential between staff and researchers, this data would need to be shared with managers. This would require moving from a position of implicit distrust to one of explicit trust in sharing data. To build this trust, managers could provide positive debriefs and tailored training based on existing data sources. Building trust is needed now to realise the full benefits from physiological data use in future.

Staff queried how accurate physiological data would be for measuring MWL. Participants understood that stress increases heart rate but queried whether it was clear what was good or bad levels in terms of workload. Responses imply signallers are more familiar and accepting of measures that indicate task demand and performance outcome, rather than measures such as wearables that indicate individual experience of MWL. Choosing to pursue wearables in the railway industry would depend on whether questions around MWL changed to considering individual experience. Further research is needed to better understand how specific aspects of physiological data interact with individual MWL experience. This would provide clarity of when it would be appropriate to measure physiological state, including alertness and stress, when measuring MWL.

Conclusions

This study found physiological wearable devices could, in future, suit use in live operations to collect physiological data. The wrist strap, with high comfort ratings and a low risk of distraction, could suit live operations. In addition, the stress and increased alertness inferred from EDA were both identified as relevant to signallers. The chest strap, whilst less familiar to staff, could suit live operations after a period of familiarisation. Further research is needed to investigate which aspects of MWL can be inferred from HRV, and the accuracy of physiological data for MWL measurement. Compared to the wearable measures, the app on a mobile device was the least suitable for live operations as it was deemed too distracting, particularly for trainees.

Despite the potential for physiological data to provide visibility of varying individual MWL, low social acceptance was predicted amongst colleagues and measures were not viewed as a status symbol. This could be explained by staff concerns being less about the devices and more around the consequences of data use, including data privacy and data misuse. Trust was a key factor, with concerns that data may be misinterpreted by managers, or staff unfairly compared, or staff losing their employment. To address these concerns, staff would prefer their data to be anonymised. Whilst anonymity would increase staff acceptance, it would be difficult to maintain data privacy in a work setting as individual identities may be inferred by staff from the roster. Instead, prior to introducing wearables, explicit trust in sharing data should be built now. This could be achieved through support for staff such as flexible breaks, positive debriefs, or tailored training. If tailored individual support provided tangible operational benefits, then physiological data from wearables could provide additional information for the collaborative management of MWL.

Acknowledgements

This research was funded by the Horizon Centre for Doctoral Training, University of Nottingham (UKRI Grant No. EP/L015463/1) and a Rail Consortium (Rail Safety Standards Board, Network Rail and Rail Delivery Group). Thank you to the Network Rail staff and participants whose support ensured the study was completed whilst COVD-19 restrictions were in place.

References

Broekhoven, R., Siefel, A.W., Schraagen, J.M.C., & Noordzij, M.L. (2016). 'Comparison of real-time relative workload measurements in rail signallers', in Rail Human Factors Proceedings of the 2nd Germany Workshop March, 8th and 9th, 2016, Stadthalle Braunschweig, pp. 30–40. Available at: http://doc.utwente.nl/99353/.

Charles, R. L. & Nixon, J. (2019). 'Measuring mental workload using physiological measures: A systematic review', Applied Ergonomics, pp. 221–232. doi: 10.1016/j.apergo.2018.08.028.

Crowley, K. & Balfe, N. (2018). Investigation of Train Driver Physiological Responses, in Proceedings of 32nd Int. BCS HCI Conference. BCS Learning & Development Ltd., pp.1–9.

Fowler, A. (2023). Human Performance in Rail: determining the potential of physiological data from wearable technologies. PhD Thesis, University of Nottingham.

Gribel, L., Regier, S., & Stengel, I. (2016). Acceptance Factors of Wearable Computing: An Empirical Investigation. In INC (pp. 67-72).

Healey, J.A. & Picard, R.W. (2005). Detecting stress during real-world driving tasks using physiological sensors. IEEE Tranactions on Intelligent Transportation Systems, 6(2), pp.156–166. doi.org/10.1109/TITS.2005.848368

Jacobs, J.V., Hettinger, L.J., Huang, Y.H., Jeffries, S., Lesch, M.F., Simmons, L.A., Verma, S.K., & Willetts, J.L. (2019). Employee acceptance of wearable technology in the workplace. Applied ergonomics, 78, pp.148-156.

Lehrer, P, Karavidas, M., Lu, S., Vaschillo, E., Vaschillo, B., & Cheng, A. (2010). Cardiac data increase association between self-report and both expert ratings of task load and task performance in flight simulator tasks: An exploratory study. International Journal of Psychophysiology, 76, pp.80–87.

Parasuraman, A. & Colby, C.L. (2015). An updated and streamlined technology readiness index: TRI 2.0. Journal of service research, 18(1), pp.59-74.

Pickup, L., Wilson, J.R., Norris, B.J., Mitchell, L., & Morrisroe, G., (2005). 'The Integrated Workload Scale (IWS): A new self-report tool to assess railway signaller workload', Applied Ergonomics, 36(6 SPEC. ISS.), pp. 681–693. doi: 10.1016/j.apergo.2005.05.004.

Saldaña, J. (2016). 'The Coding Manual for Qualitative Researchers (No. 14)', Sage.

Sharples, S. (2019). 'Workload II: A future paradigm for analysis and measurement', in Advances in Intelligent Systems and Computing, pp. 489–498. doi: 10.1007/978-3-319-96071-5_52.

Song, Y.S., Baek, J.H., Hwang, D.S., Lee, J.W., Lee, Y.J., Park, H.J., Choi, J.H., & Yang, H.K. (2014). 'Physiological Status Assessment of Locomotive Engineer During Train Operation', Journal of Electrical, 9(1), pp.324-333.

Tortora, G. & Derrickson, B. (2007). Introduction to the Human Body. Wiley.

Urquhart, L. & Craigon, P. (2020). The Moral-IT Deck: A Tool for Ethics by Design. arXiv preprint arXiv:2007.07514.

Venkatesh, V. & Davis, F.D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. Management science, 46(2), pp.186-204.

Wolf, P., Menzel, F., & Rennhak, C. (2018). An Extension of the Technology Acceptance Model tailored to Wearable Device Technology. Munich Business School Working Paper Series.